

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

WO 92/06176

(51) International Patent Classification 5 :	C12N 1/24, 15/00, C07H 21/00	(11) International Publication Number:	WO 92/06176
	A1	(43) International Publication Date:	16 April 1992 (15.04.92)

(21) International Application Number:	PCT/US91/07141	(81) Designated States:	AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH (European patent), CI (OAPI patent), CM (OAPI patent), CS, DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GA (OAPI patent), GB (European patent), GN (OAPI patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU (European patent), MC, MG, ML (OAPI patent), MN, MR (OAPI patent), MW, NL (European patent), NO, PL, RO, SD, SE (European patent), SN (OAPI patent), SU, TD (OAPI patent), TG (OAPI patent).
(22) International Filing Date:	27 September 1991 (27.09.91)		
(30) Priority data:	590,664	28 September 1990 (28.09.90); US	
(71) Applicant:	IXSYS, INC. [US/US]; 3550 General Atomics Court, Suite L103, San Diego, CA 92121 (US).		
(72) Inventor:	HUSE, William, D. ; 471 Avenida Primavera, Del Mar, CA 92014 (US).		
(74) Agents:	CAMPBELL, Cathryn et al.; Pretty, Schroeder, Brueggemann & Clark, 444 South Flower Street, Suite 2000, Los Angeles, CA 90071 (US).	Published	<i>With international search report</i>

(54) Title: SURFACE EXPRESSION LIBRARIES OF RANDOMIZED PEPTIDES

(57) Abstract

A composition of matter comprising a plurality of prokaryotic cells containing a diverse population of expressible oligonucleotides operationally linked to expression elements, said expressible oligonucleotides having a desirable bias of random codon sequences.

+ DESIGNATIONS OF "SU"

Any designation of "SU" has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CP	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU+	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TC	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark				

SURFACE EXPRESSION LIBRARIES
OF RANDOMIZED PEPTIDES

5

BACKGROUND OF THE INVENTION

This invention relates generally to methods for synthesizing and expressing oligonucleotides and, more particularly, to methods for expressing oligonucleotides having random codon sequences.

10 Oligonucleotide synthesis proceeds via linear coupling of individual monomers in a stepwise reaction. The reactions are generally performed on a solid phase support by first coupling the 3' end of the first monomer to the support. The second monomer is added to the 5' end of the 15 first monomer in a condensation reaction to yield a dinucleotide coupled to the solid support. At the end of each coupling reaction, the by-products and unreacted, free monomers are washed away so that the starting material for the next round of synthesis is the pure oligonucleotide attached to the support. In this reaction scheme, the 20 stepwise addition of individual monomers to a single, growing end of a oligonucleotide ensures accurate synthesis of the desired sequence. Moreover, unwanted side reactions are eliminated, such as the condensation of two 25 oligonucleotides, resulting in high product yields.

In some instances, it is desired that synthetic oligonucleotides have random nucleotide sequences. This result can be accomplished by adding equal proportions of all four nucleotides in the monomer coupling reactions, 30 leading to the random incorporation of all nucleotides and yielding a population of oligonucleotides with random sequences. Since all possible combinations of nucleotide sequences are represented within the population, all possible codon triplets will also be represented. If the

objective is ultimately to generate random peptide products, this approach has a severe limitation because the random codons synthesized will bias the amino acids incorporated during translation of the DNA by the cell into 5 polypeptides.

The bias is due to the redundancy of the genetic code. There are four nucleotide monomers which leads to sixty-four possible triplet codons. With only twenty amino acids to specify, many of the amino acids are encoded by multiple 10 codons. Therefore, a population of oligonucleotides synthesized by sequential addition of monomers from a random population will not encode peptides whose amino acid sequence represents all possible combinations of the twenty different amino acids in equal proportions. That is, the 15 frequency of amino acids incorporated into polypeptides will be biased toward those amino acids which are specified by multiple codons.

To alleviate amino acid bias due to the redundancy of the genetic code, the oligonucleotides can be synthesized 20 from nucleotide triplets. Here, a triplet coding for each of the twenty amino acids is synthesized from individual monomers. Once synthesized, the triplets are used in the coupling reactions instead of individual monomers. By 25 mixing equal proportions of the triplets, synthesis of oligonucleotides with random codons can be accomplished. However, the cost of synthesis from such triplets far exceeds that of synthesis from individual monomers because triplets are not commercially available.

Amino acid bias can be reduced, however, by 30 synthesizing the degenerate codon sequence NNK where N is a mixture of all four nucleotides and K is a mixture of guanine and thymine nucleotides. Each position within an oligonucleotide having this codon sequence will contain a total of 32 codons (12 encoding amino acids being

represented once, 5 represented twice, 3 represented three times and one codon being a stop codon). Oligonucleotides expressed with such degenerate codon sequences will produce peptide products whose sequences are biased toward those 5 amino acids being represented more than once. Thus, populations of peptides whose sequences are completely random cannot be obtained from oligonucleotides synthesized from degenerate sequences.

There thus exists a need for a method to express 10 oligonucleotides having a fully random or desirably biased sequence which alleviates genetic redundancy. The present invention satisfies these needs and provides additional advantages as well.

SUMMARY OF THE INVENTION

15 The invention provides a plurality of prokaryotic cells containing a diverse population of expressible oligonucleotides operationally linked to expression elements, the expressible oligonucleotides having a desirable bias of random codon sequences.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic drawing for synthesizing oligonucleotides from nucleotide monomers with random triplets at each position using twenty reaction vessels.

20 Figure 2 is a schematic drawing for synthesizing oligonucleotides from nucleotide monomers with random triplets at each position using ten reaction vessels.

Figure 3 is a schematic diagram of the two vectors used for sublibrary and library production from precursor oligonucleotide portions. M13IX22 (Figure 3A) is the 30 vector used to clone the anti-sense precursor portions

(hatched box). The single-headed arrow represents the Lac p/o expression sequences and the double-headed arrow represents the portion of M13IX22 which is to be combined with M13IX42. The amber stop codon for biological selection and relevant restriction sites are also shown. 5 M13IX42 (Figure 3B) is the vector used to clone the sense precursor portions (open box). Thick lines represent the pseudo-wild type (ψ gVIII) and wild type (gVIII) gene VIII sequences. The double-headed arrow represents the portion 10 of M13IX42 which is to be combined with M13IX22. The two amber stop codons and relevant restriction sites are also shown. Figure 3C shows the joining of vector population from sublibraries to form the functional surface expression 15 vector M13IX. Figure 3D shows the generation of a surface expression library in a non-suppressor strain and the production of phage. The phage are used to infect a suppressor strain (Figure 3E) for surface expression and screening of the library.

Figure 4 is a schematic diagram of the vector used for 20 generation of surface expression libraries from random oligonucleotide populations (M13IX30). The symbols are as described for Figure 3.

Figure 5 is the nucleotide sequence of M13IX42 (SEQ ID NO: 1).

25 Figure 6 is the nucleotide sequence of M13IX22 (SEQ ID NO: 2).

Figure 7 is the nucleotide sequence of M13IX30 (SEQ ID NO: 3).

30 Figure 8 is the nucleotide sequence of M13ED03 (SEQ ID NO: 4).

Figure 9 is the nucleotide sequence of M13IX421 (SEQ

ID NO: 5).

Figure 10 is the nucleotide sequence of M13ED04 (SEQ ID NO: 6).

DETAILED DESCRIPTION OF THE INVENTION

5 This invention is directed to a simple and inexpensive method for synthesizing and expressing oligonucleotides having a desirable bias of random codons using individual monomers. The method is advantageous in that individual monomers are used instead of triplets and by synthesizing 10 only a non-degenerate subset of all triplets, codon redundancy is alleviated. Thus, the oligonucleotides synthesized represent a large proportion of possible random triplet sequences which can be obtained. The oligonucleotides can be expressed, for example, on the 15 surface of filamentous bacteriophage in a form which does not alter phage viability or impose biological selections against certain peptide sequences. The oligonucleotides produced are therefore useful for generating an unlimited number of pharmacological and research products.

20 In one embodiment, the invention entails the sequential coupling of monomers to produce oligonucleotides with a desirable bias of random codons. The coupling reactions for the randomization of twenty codons which specify the amino acids of the genetic code are performed 25 in ten different reaction vessels. Each reaction vessel contains a support on which the monomers for two different codons are coupled in three sequential reactions. One of the reactions couples an equal mixture of two monomers such that the final product has two different codon sequences. 30 The codons are randomized by removing the supports from the reaction vessels and mixing them to produce a single batch of supports containing all twenty codons at a particular position. Synthesis at the next codon position proceeds by

equally dividing the mixed batch of supports into ten reaction vessels as before and sequentially coupling the monomers for each pair of codons. The supports are again mixed to randomize the codons at the position just synthesized. The cycle of coupling, mixing and dividing continues until the desired number of codon positions have been randomized. After the last position has been randomized, the oligonucleotides with random codons are cleaved from the support. The random oligonucleotides can then be expressed, for example, on the surface of filamentous bacteriophage as gene VIII-peptide fusion proteins. Alternative genes can be used as well.

In its broadest form, the invention provides a diverse population of synthetic oligonucleotides contained in vectors so as to be expressible in cells. Such populations of diverse oligonucleotides can be fully random at one or more codon sites or can be fully defined at one or more sites, so long as at least one site the codons are randomly variable. The populations of oligonucleotides can be expressed as fusion products in combination with surface proteins of filamentous bacteriophage, such as M13, as with gene VIII. The vectors can be transfected into a plurality of cells, such as the prokaryote E. coli.

The diverse population of oligonucleotides can be formed by randomly combining first and second precursor populations, each precursor population having a desirable bias of random codon sequences. Methods of synthesizing and expressing the diverse population of expressible oligonucleotides are also provided.

In a preferred embodiment, two populations of random oligonucleotides are synthesized. The oligonucleotides within each population encode a portion of the final oligonucleotide which is to be expressed. Oligonucleotides within one population encode the carboxy terminal portion

of the expressed oligonucleotides. These oligonucleotides are cloned in frame with a gene VIII (gVIII) sequence so that translation of the sequence produces peptide fusion proteins. The second population of oligonucleotides are 5 cloned into a separate vector. Each oligonucleotide within this population encodes the anti-sense of the amino terminal portion of the expressed oligonucleotides. This vector also contains the elements necessary for expression. The two vectors containing the random oligonucleotides are 10 combined such that the two precursor oligonucleotide portions are joined together at random to form a population of larger oligonucleotides derived from two smaller portions. The vectors contain selectable markers to ensure maximum efficiency in joining together the two 15 oligonucleotide populations. A mechanism also exists to control the expression of gVIII-peptide fusion proteins during library construction and screening.

As used herein, the term "monomer" or "nucleotide monomer" refers to individual nucleotides used in the 20 chemical synthesis of oligonucleotides. Monomers that can be used include both the ribo- and deoxyribo- forms of each of the five standard nucleotides (derived from the bases adenine (A or dA, respectively), guanine (G or dG), cytosine (C or dC), thymine (T) and uracil (U)). 25 Derivatives and precursors of bases such as inosine which are capable of supporting polypeptide biosynthesis are also included as monomers. Also included are chemically modified nucleotides, for example, one having a reversible blocking agent attached to any of the positions on the 30 purine or pyrimidine bases, the ribose or deoxyribose sugar or the phosphate or hydroxyl moieties of the monomer. Such blocking groups include, for example, dimethoxytrityl, benzoyl, isobutyryl, beta-cyanoethyl and diisopropylamine groups, and are used to protect hydroxyls, exocyclic amines 35 and phosphate moieties. Other blocking agents can also be used and are known to one skilled in the art.

As used herein, the term "tuple" refers to a group of elements of a definable size. The elements of a tuple as used herein are nucleotide monomers. For example, a tuple can be a dinucleotide, a trinucleotide or can also be four 5 or more nucleotides.

As used herein, the term "codon" or "triplet" refers to a tuple consisting of three adjacent nucleotide monomers which specify one of the twenty naturally occurring amino acids found in polypeptide biosynthesis. 10 The term also includes nonsense, or stop, codons which do not specify any amino acid.

"Random codons" or "randomized codons," as used herein, refers to more than one codon at a position within a collection of oligonucleotides. The number of different 15 codons can be from two to twenty at any particular position. "Randomized oligonucleotides," as used herein, refers to a collection of oligonucleotides with random codons at one or more positions. "Random codon sequences" as used herein means that more than one codon position 20 within a randomized oligonucleotide contains random codons. For example, if randomized oligonucleotides are six nucleotides in length (i.e., two codons) and both the first and second codon positions are randomized to encode all twenty amino acids, then a population of oligonucleotides 25 having random codon sequences with every possible combination of the twenty triplets in the first and second position makes up the above population of randomized oligonucleotides. The number of possible codon combinations is 20^2 . Likewise, if randomized oligonucleotides of fifteen nucleotides in length are 30 synthesized which have random codon sequences at all positions encoding all twenty amino acids, then all triplets coding for each of the twenty amino acids will be found in equal proportions at every position. The 35 population constituting the randomized oligonucleotides

will contain 20^{15} different possible species of oligonucleotides. "Random triplets," or "randomized triplets" are defined analogously.

As used herein, the term "bias" refers to a 5 preference. It is understood that there can be degrees of preference or bias toward codon sequences which encode particular amino acids. For example, an oligonucleotide whose codon sequences do not preferably encode particular amino acids is unbiased and therefore completely random. 10 The oligonucleotide codon sequences can also be biased toward predetermined codon sequences or codon frequencies and while still diverse and random, will exhibit codon sequences biased toward a defined, or preferred, sequence. 15 "A desirable bias of random codon sequences" as used herein, refers to the predetermined degree of bias which can be selected from totally random to essentially, but not totally, defined (or preferred). There must be at least one codon position which is variable, however.

As used herein, the term "support" refers to a solid 20 phase material for attaching monomers for chemical synthesis. Such support is usually composed of materials such as beads of control pore glass but can be other materials known to one skilled in the art. The term is also meant to include one or more monomers coupled to the 25 support for additional oligonucleotide synthesis reactions.

As used herein, the terms "coupling" or "condensing" refers to the chemical reactions for attaching one monomer to a second monomer or to a solid support. Such reactions are known to one skilled in the art and are typically 30 performed on an automated DNA synthesizer such as a MilliGen/Bioscience Cyclone Plus Synthesizer using procedures recommended by the manufacturer. "Sequentially coupling" as used herein, refers to the stepwise addition of monomers.

A method of synthesizing oligonucleotides having random triplets using individual monomers is described. The method consists of several steps, the first being synthesis of a nucleotide triplet for each triplet to be randomized.

5 As described here and below, a nucleotide triplet (i.e., a codon) will be used as a specific example of a triplet. Any size triplet will work using the methods disclosed herein, and one skilled in the art would know how to use the methods to randomize triplets of any size.

10 If the randomization of codons specifying all twenty amino acids is desired at a position, then twenty different codons are synthesized. Likewise, if randomization of only ten codons at a particular position is desired then those ten codons are synthesized. Randomization of codons from

15 two to sixty-four can be accomplished by synthesizing each desired triplet. Preferably, randomization of from two to twenty codons is used for any one position because of the redundancy of the genetic code. The codons selected at one position do not have to be the same codons selected at the

20 next position. Additionally, the sense or anti-sense sequence oligonucleotide can be synthesized. The process therefore provides for randomization of any desired codon position with any number of codons.

Codons to be randomized are synthesized sequentially by coupling the first monomer of each codon to separate supports. The supports for the synthesis of each codon can, for example, be contained in different reaction vessels such that one reaction vessel corresponds to the monomer coupling reactions for one codon. As will be used

25 here and below, if twenty codons are to be randomized, then twenty reaction vessels can be used in independent coupling reactions for the first twenty monomers of each codon. Synthesis proceeds by sequentially coupling the second monomer of each codon to the first monomer to produce a

30 dimer, followed by coupling the third monomer for each

codon to each of the above-synthesized dimers to produce a trimer (Figure 1, step 1, where M_1 , M_2 and M_3 represent the first, second and third monomer, respectively, for each codon to be randomized).

5 Following synthesis of the first codons from individual monomers, the randomization is achieved by mixing the supports from all twenty reaction vessels which contain the individual codons to be randomized. The solid phase support can be removed from its vessel and mixed to 10 achieve a random distribution of all codon species within the population (Figure 1, step 2). The mixed population of supports, constituting all codon species, are then redistributed into twenty independent reaction vessels 15 (Figure 1, step 3). The resultant vessels are all identical and contain equal portions of all twenty codons coupled to a solid phase support.

For randomization of the second position codon, synthesis of twenty additional codons is performed in each of the twenty reaction vessels produced in step 3 as the condensing substrates of step 1 (Figure 1, step 4). Steps 20 1 and 4 are therefore equivalent except that step 4 uses the supports produced by the previous synthesis cycle (steps 1 through 3) for codon synthesis whereas step 1 is the initial synthesis of the first codon in the 25 oligonucleotide. The supports resulting from step 4 will each have two codons attached to them (i.e., a hexanucleotide) with the codon at the first position being any one of twenty possible codons (i.e., random) and the codon at the second position being one of the twenty 30 possible codons.

For randomization of the codon at the second position and synthesis of the third position codon, steps 2 through 4 are again repeated. This process yields in each vessel a three codon oligonucleotide (i.e., 9 nucleotides) with

codon positions 1 and 2 randomized and position three containing one of the twenty possible codons. Steps 2 through 4 are repeated to randomize the third position codon and synthesize the codon at the next position. The 5 process is continued until an oligonucleotide of the desired length is achieved. After the final randomization step, the oligonucleotide can be cleaved from the supports and isolated by methods known to one skilled in the art. Alternatively, the oligonucleotides can remain on the 10 supports for use in methods employing probe hybridization.

The diversity of codon sequences, i.e., the number of different possible oligonucleotides, which can be obtained using the methods of the present invention, is extremely large and only limited by the physical characteristics of 15 available materials. For example, a support composed of beads of about 100 μm in diameter will be limited to about 10,000 beads/reaction vessel using a 1 μM reaction vessel containing 25 mg of beads. This size bead can support about 1×10^7 oligonucleotides per bead. Synthesis using 20 separate reaction vessels for each of the twenty amino acids will produce beads in which all the oligonucleotides attached to an individual bead are identical. The diversity which can be obtained under these conditions is approximately 10^7 copies of $10,000 \times 20$ or 200,000 different 25 random oligonucleotides. The diversity can be increased, however, in several ways without departing from the basic methods disclosed herein. For example, the number of possible sequences can be increased by decreasing the size of the individual beads which make up the support. A bead 30 of about 30 μm in diameter will increase the number of beads per reaction vessel and therefore the number of oligonucleotides synthesized. Another way to increase the diversity of oligonucleotides with random codons is to increase the volume of the reaction vessel. For example, 35 using the same size bead, a larger volume can contain a greater number of beads than a smaller vessel and therefore

support the synthesis of a greater number of oligonucleotides. Increasing the number of codons coupled to a support in a single reaction vessel also increases the diversity of the random oligonucleotides. The total diversity will be the number of codons coupled per vessel raised to the number of codon positions synthesized. For example, using ten reaction vessels, each synthesizing two codons to randomize a total of twenty codons, the number of different oligonucleotides of ten codons in length per 100 μm bead can be increased where each bead will contain about 2^{10} or 1×10^3 different sequences instead of one. One skilled in the art will know how to modify such parameters to increase the diversity of oligonucleotides with random codons.

A method of synthesizing oligonucleotides having random codons at each position using individual monomers wherein the number of reaction vessels is less than the number of codons to be randomized is also described. For example, if twenty codons are to be randomized at each position within an oligonucleotide population, then ten reaction vessels can be used. The use of a smaller number of reaction vessels than the number of codons to be randomized at each position is preferred because the smaller number of reaction vessels is easier to manipulate and results in a greater number of possible oligonucleotides synthesized.

The use of a smaller number of reaction vessels for random synthesis of twenty codons at a desired position within an oligonucleotide is similar to that described above using twenty reaction vessels except that each reaction vessel can contain the synthesis products of more than one codon. For example, step one synthesis using ten reaction vessels proceeds by coupling about two different codons on supports contained in each of ten reaction vessels. This is shown in Figure 2 where each of the two

codons coupled to a different support can consist of the following sequences: (1) (T/G)TT for Phe and Val; (2) (T/C)CT for Ser and Pro; (3) (T/C)AT for Tyr and His; (4) (T/C)GT for Cys and Arg; (5) (C/A)TG for Leu and Met; (6) (C/G)AG for Gln and Glu; (7) (A/G)CT for Thr and Ala; (8) (A/G)AT for Asn and Asp; (9) (T/G)GG for Trp and Gly and (10) A(T/A)A for Ile and Cys. The slash (/) signifies that a mixture of the monomers indicated on each side of the slash are used as if they were a single monomer in the indicated coupling step. The antisense sequence for each of the above codons can be generated by synthesizing the complementary sequence. For example, the antisense for Phe and Val can be AA(C/A). The amino acids encoded by each of the above pairs of sequences are given as the standard three letter nomenclature.

Coupling of the monomers in this fashion will yield codons specifying all twenty of the naturally occurring amino acids attached to supports in ten reaction vessels. However, the number of individual reaction vessels to be used will depend on the number of codons to be randomized at the desired position and can be determined by one skilled in the art. For example, if ten codons are to be randomized, then five reaction vessels can be used for coupling. The codon sequences given above can be used for this synthesis as well. The sequences of the codons can also be changed to incorporate or be replaced by any of the additional forty-four codons which constitutes the genetic code.

The remaining steps of synthesis of oligonucleotides with random codons using a smaller number of reaction vessels are as outlined above for synthesis with twenty reaction vessels except that the mixing and dividing steps are performed with supports from about half the number of reaction vessels. These remaining steps are shown in Figure 2 (steps 2 through 4).

Oligonucleotides having at least one specified tuplet at a predetermined position and the remaining positions having random triplets can also be synthesized using the methods described herein. The synthesis steps are similar to those outlined above using twenty or less reaction vessels except that prior to synthesis of the specified codon position, the dividing of the supports into separate reaction vessels for synthesis of different codons is omitted. For example, if the codon at the second position of the oligonucleotide is to be specified, then following synthesis of random codons at the first position and mixing of the supports, the mixed supports are not divided into new reaction vessels but, instead, can be contained in a single reaction vessel to synthesize the specified codon. The specified codon is synthesized sequentially from individual monomers as described above. Thus, the number of reaction vessels can be increased or decreased at each step to allow for the synthesis of a specified codon or a desired number of random codons.

Following codon synthesis, the mixed supports are divided into individual reaction vessels for synthesis of the next codon to be randomized (Figure 1, step 3) or can be used without separation for synthesis of a consecutive specified codon. The rounds of synthesis can be repeated for each codon to be added until the desired number of positions with predetermined or randomized codons are obtained.

Synthesis of oligonucleotides with the first position codon being specified can also be synthesized using the above method. In this case, the first position codon is synthesized from the appropriate monomers. The supports are divided into the required number of reaction vessels needed for synthesis of random codons at the second position and the rounds of synthesis, mixing and dividing are performed as described above.

A method of synthesizing oligonucleotides having triplets which are diverse but biased toward a predetermined sequence is also described herein. This method employs two reaction vessels, one vessel for the synthesis of a predetermined sequence and the second vessel for the synthesis of a random sequence. This method is advantageous to use when a significant number of codon positions, for example, are to be of a specified sequence since it alleviates the use of multiple reaction vessels. Instead, a mixture of four different monomers such as adenine, guanine, cytosine and thymine nucleotides are used for the first and second monomers in the codon. The codon is completed by coupling a mixture of a pair of monomers of either guanine and thymine or cytosine and adenine nucleotides at the third monomer position. In the second vessel, nucleotide monomers are coupled sequentially to yield the predetermined codon sequence. Mixing of the two supports yields a population of oligonucleotides containing both the predetermined codon and the random codons at the desired position. Synthesis can proceed by using this mixture of supports in a single reaction vessel, for example, for coupling additional predetermined codons or, further dividing the mixture into two reaction vessels for synthesis of additional random codons.

The two reaction vessel method can be used for codon synthesis within an oligonucleotide with a predetermined triplet sequence by dividing the support mixture into two portions at the desired codon position to be randomized. Additionally, this method allows for the extent of randomization to be adjusted. For example, unequal mixing or dividing of the two supports will change the fraction of codons with predetermined sequences compared to those with random codons at the desired position. Unequal mixing and dividing of supports can be useful when there is a need to synthesize random codons at a significant number of positions within an oligonucleotide of a longer or shorter

length.

The extent of randomization can also be adjusted by using unequal mixtures of monomers in the first, second and third monomer coupling steps of the random codon position.

5 The unequal mixtures can be in any or all of the coupling steps to yield a population of codons enriched in sequences reflective of the monomer proportions.

Synthesis of randomized oligonucleotides is performed using methods well known to one skilled in the art. Linear coupling of monomers can, for example, be accomplished using phosphoramidite chemistry with a MilliGen/Bioscience Cyclone Plus automated synthesizer as described by the manufacturer (Millipore, Burlington, MA). Other chemistries and automated synthesizers can be employed as well and are known to one skilled in the art.

Synthesis of multiple codons can be performed without modification to the synthesizer by separately synthesizing the codons in individual sets of reactions. Alternatively, modification of an automated DNA synthesizer can be performed for the simultaneous synthesis of codons in multiple reaction vessels.

In one embodiment, the invention provides a plurality of prokaryotic cells containing a diverse population of expressible oligonucleotides operationally linked to expression elements, the expressible oligonucleotides having a desirable bias of random codon sequences produced from diverse combinations of first and second oligonucleotides having a desirable bias of random sequences. The invention provides for a method for 25 constructing such a plurality of prokaryotic cells as well.

The oligonucleotides synthesized by the above methods can be used to express a plurality of random peptides which

are unbiased, diverse but biased toward a predetermined sequence or which contain at least one specified codon at a predetermined position. The need will determine which type of oligonucleotide is to be expressed to give the 5 resultant population of random peptides and is known to one skilled in the art. Expression can be performed in any compatible vector/host system. Such systems include, for example, plasmids or phagemids in prokaryotes such as *E. coli*, yeast systems, and other eucaryotic systems such as 10 mammalian cells, but will be described herein in context with its presently preferred embodiment, i.e. expression on the surface of filamentous bacteriophage. Filamentous bacteriophage can be, for example, M13, f1 and fd. Such 15 phage have circular single-stranded genomes and double strand replicative DNA forms. Additionally, the peptides can also be expressed in soluble or secreted form depending on the need and the vector/host system employed.

Expression of random peptides on the surface of M13 can be accomplished, for example, using the vector system 20 shown in Figure 3. Construction of the vectors enabling one of ordinary skill to make them are explicitly set out in Examples I and II. The complete nucleotide sequences are given in Figures 5, 6 and 7 (SEQ ID NOS: 1, 2 and 3, respectively). This system produces random 25 oligonucleotides functionally linked to expression elements and to gVIII by combining two smaller oligonucleotide portions contained in separate vectors into a single vector. The diversity of oligonucleotide species obtained by this system or others described herein can be 5×10^7 or 30 greater. Diversity of less than 5×10^7 can also be obtained and will be determined by the need and type of random peptides to be expressed. The random combination of two precursor portions into a larger oligonucleotide increases the diversity of the population several fold and 35 has the added advantage of producing oligonucleotides larger than what can be synthesized by standard methods.

Additionally, although the correlation is not known, when the number of possible paths an oligonucleotide can take during synthesis such as described herein is greater than the number of beads, then there will be a correlation 5 between the synthesis path and the sequences obtained. By combining oligonucleotide populations which are synthesized separately, this correlation will be destroyed. Therefore, any bias which may be inherent in the synthesis procedures will be alleviated by joining two precursor portions into 10 a contiguous random oligonucleotide.

Populations of precursor oligonucleotides to be combined into an expressible form are each cloned into separate vectors. The two precursor portions which make up the combined oligonucleotide corresponds to the carboxy and 15 amino terminal portions of the expressed peptide. Each precursor oligonucleotide can encode either the sense or anti-sense and will depend on the orientation of the expression elements and the gene encoding the fusion portion of the protein as well as the mechanism used to 20 join the two precursor oligonucleotides. For the vectors shown in Figure 3, precursor oligonucleotides corresponding to the carboxy terminal portion of the peptide encode the sense strand. Those corresponding to the amino terminal portion encode the anti-sense strand. Oligonucleotide 25 populations are inserted between the Eco RI and Sac I restriction enzyme sites in M13IX22 and M13IX42 (Figure 3A and B). M13IX42 (SEQ ID NO: 1) is the vector used for sense strand precursor oligonucleotide portions and M13IX22 (SEQ ID NO: 2) is used for anti-sense precursor portions.

30 The populations of randomized oligonucleotides inserted into the vectors are synthesized with Eco RI and Sac I recognition sequences flanking opposite ends of the random codon sequences. The sites allow annealing and ligation of these single strand oligonucleotides into a 35 double stranded vector restricted with Eco RI and Sac I.

Alternatively, the oligonucleotides can be inserted into the vector by standard mutagenesis methods. In this latter method, single stranded vector DNA is isolated from the phage and annealed with random oligonucleotides having known sequences complementary to vector sequences. The oligonucleotides are extended with DNA polymerase to produce double stranded vectors containing the randomized oligonucleotides.

The vector used for sense strand oligonucleotide portions, M13IX42 (Figure 3B) contains down-stream and in frame with the Eco RI and Sac I restriction sites a sequence encoding the pseudo-wild type gVIII product. This gene encodes the wild type M13 gVIII amino acid sequence but has been changed at the nucleotide level to reduce homologous recombination with the wild type gVIII contained on the same vector. The wild type gVIII is present to ensure that at least some functional, non-fusion coat protein will be produced. The inclusion of a wild type gVIII therefore reduces the possibility of non-viable phage production and biological selection against certain peptide fusion proteins. Differential regulation of the two genes can also be used to control the relative ratio of the pseudo and wild type proteins.

Also contained downstream and in frame with the Eco RI and Sac I restriction sites is an amber stop codon. The mutation is located six codons downstream from Sac I and therefore lies between the inserted oligonucleotides and the gVIII sequence. As was the function of the wild type gVIII, the amber stop codon also reduces biological selection when combining precursor portions to produce expressible oligonucleotides. This is accomplished by using a non-suppressor ($sup\ 0$) host strain because non-suppressor strains will terminate expression after the oligonucleotide sequences but before the pseudo gVIII sequences. Therefore, the pseudo gVIII will never be

expressed on the phage surface under these circumstances. Instead, only soluble peptides will be produced. Expression in a non-suppressor strain can be advantageously utilized when one wishes to produce large populations of soluble peptides. Stop codons other than amber, such as 5 opal and ochre, or molecular switches, such as inducible repressor elements, can also be used to unlink peptide expression from surface expression. Additional controls exist as well and are described below.

10 The vector used for anti-sense strand oligonucleotide portions, M13IX22, (Figure 3A), contains the expression elements for the peptide fusion proteins. Upstream and in frame with the Sac I and Eco RI sites in this vector is a leader sequence for surface expression. A ribosome binding 15 site and Lac Z promoter/operator elements are present for transcription and translation of the peptide fusion proteins.

Both vectors contain a pair of Fok I restriction 20 enzyme sites (Figure 3 A and B) for joining together two precursor oligonucleotide portions and their vector sequences. One site is located at the ends of each precursor oligonucleotide which is to be joined. The second Fok I site within the vectors is located at the end 25 of the vector sequences which are to be joined. The 5' overhang of this second Fok I site has been altered to encode a sequence which is not found in the overhangs produced at the first Fok I site within the oligonucleotide portions. The two sites allow the cleavage of each 30 circular vector into two portions and subsequent ligation of essential components within each vector into a single circular vector where the two oligonucleotide precursor portions form a contiguous sequence (Figure 3C). Non-compatible overhangs produced at the two Fok I sites allows 35 optimal conditions to be selected for performing concatemORIZATION or circularization reactions for joining

the two vector portions. Such selection of conditions can be used to govern the reaction order and therefore increase the efficiency of joining.

Fok I is a restriction enzyme whose recognition sequence is distal to the point of cleavage. Distal placement of the recognition sequence in its location to the cleavage point is important since if the two were superimposed within the oligonucleotide portions to be combined, it would lead to an invariant codon sequence at the juncture. To alleviate the formation of invariant codons at the juncture, Fok I recognition sequences can be placed outside of the random codon sequence and still be used to restrict within the random sequence. Subsequent annealing of the single-strand overhangs produced by Fok I and ligation of the two oligonucleotide precursor portions allows the juncture to be formed. A variety of restriction enzymes restrict DNA by this mechanism and can be used instead of Fok I to join precursor oligonucleotides without creating invariant codon sequences. Such enzymes include, for example, Alw I, Bbv I, Bsp MI, Hga I, Hph I, Mbo II, Mnl I, Pst I and Sfa NI. One skilled in the art knows how to substitute Fok I recognition sequences for alternative enzyme recognition sequences such as those above, and use the appropriate enzyme for joining precursor oligonucleotide portions.

Although the sequences of the precursor oligonucleotides are random and will invariably have oligonucleotides within the two precursor populations whose sequences are sufficiently complementary to anneal after cleavage, the efficiency of annealing can be increased by insuring that the single-strand overhangs within one precursor population will have a complementary sequence within the second precursor population. This can be accomplished by synthesizing a non-degenerate series of known sequences at the Fok I cleavage site coding for each

of the twenty amino acids. Since the Fok I cleavage site contains a four base overhang, forty different sequences are needed to randomly encode all twenty amino acids. For example, if two precursor populations of ten codons in 5 length are to be combined, then after the ninth codon position is synthesized, the mixed population of supports are divided into forty reaction vessels for each of the populations and complementary sequences for each of the corresponding reaction vessels between populations are 10 independently synthesized. The sequences are shown in Tables III and VI of Example I where the oligonucleotides on columns 1R through 40R form complementary overhangs with the oligonucleotides on the corresponding columns 1L through 40L once cleaved. The degenerate X positions in 15 Table VI are necessary to maintain the reading frame once the precursor oligonucleotide portions are joined. However, use of restriction enzymes which produce a blunt end, such as Mnl I can be alternatively used in place of Fok I to alleviate the degeneracy introduced in maintaining 20 the reading frame.

The last feature exhibited by each of the vectors is an amber stop codon located in an essential coding sequence within the vector portion lost during combining (Figure 3C). The amber stop codon is present to select for viable 25 phage produced from only the proper combination of precursor oligonucleotides and their vector sequences into a single vector species. Other non-sense mutations or selectable markers can work as well.

The combining step randomly brings together different 30 precursor oligonucleotides within the two populations into a single vector (Figure 3C; M13IX). The vector sequences donated from each independent vector, M13IX22 and M13IX42, are necessary for production of viable phage. Also, since the expression elements are contained in M13IX22 and the 35 gVIII sequences are contained in M13IX42, expression of

functional gVIII-peptide fusion proteins cannot be accomplished until the sequences are linked as shown in M13IX.

The combining step is performed by restricting each 5 population of vectors containing randomized oligonucleotides with Fok I, mixing and ligating (Figure 3C). Any vectors generated which contain an amber stop codon will not produce viable phage when introduced into a non-suppressor strain (Figure 3D). Therefore, only the 10 sequences which do not contain an amber stop codon will make up the final population of vectors contained in the library. These vector sequences are the sequences required 15 for surface expression of randomized peptides. By analogous methodology, more than two vector portions can be combined into a single vector which expresses random peptides.

The invention provides for a method of selecting peptides capable of being bound by a ligand binding protein from a population of random peptides by (a) operationally 20 linking a diverse population of first oligonucleotides having a desirable bias of random codon sequences to a first vector; (b) operationally linking a diverse population of second oligonucleotides having a desirable bias of random codon sequences to a second vector; (c) 25 combining the vector products of steps (a) and (b) under conditions where said populations of first and second oligonucleotides are joined together into a population of combined vectors; (d) introducing said population of combined vectors into a compatible host under conditions 30 sufficient for expressing said population of random peptides; and (e) determining the peptides which bind to said binding protein. The invention also provides for determining the encoding nucleic acid sequence of such peptides as well.

Surface expression of the random peptide library is performed in an amber suppressor strain. As described above, the amber stop codon between the random codon sequence and the gVIII sequence unlinks the two components 5 in a non-suppressor strain. Isolating the phage produced from the non-suppressor strain and infecting a suppressor strain will link the random codon sequences to the gVIII sequence during expression (Figure 3E). Culturing the suppressor strain after infection allows the expression of 10 all peptide species within the library as gVIII-peptide fusion proteins. Alternatively, the DNA can be isolated from the non-suppressor strain and then introduced into a suppressor strain to accomplish the same effect.

The level of expression of gVIII-peptide fusion 15 proteins can additionally be controlled at the transcriptional level. The gVIII-peptide fusion proteins are under the inducible control of the Lac Z promoter/operator system. Other inducible promoters can work as well and are known by one skilled in the art. For 20 high levels of surface expression, the suppressor library is cultured in an inducer of the Lac Z promoter such as isopropylthio- β -galactoside (IPTG). Inducible control is beneficial because biological selection against non-functional gVIII-peptide fusion proteins can be minimized 25 by culturing the library under non-expressing conditions. Expression can then be induced only at the time of screening to ensure that the entire population of oligonucleotides within the library are accurately represented on the phage surface. Also this can be used to 30 control the valency of the peptide on the phage surface.

The surface expression library is screened for 35 specific peptides which bind ligand binding proteins by standard affinity isolation procedures. Such methods include, for example, panning, affinity chromatography and solid phase blotting procedures. Panning as described by

Parmley and Smith, Gene 73:305-318 (1988), which is incorporated herein by reference, is preferred because high titers of phage can be screened easily, quickly and in small volumes. Furthermore, this procedure can select 5 minor peptide species within the population, which otherwise would have been undetectable, and amplified to substantially homogenous populations. The selected peptide sequences can be determined by sequencing the nucleic acid encoding such peptides after amplification of the phage 10 population.

The invention provides a plurality of prokaryotic cells containing a diverse population of oligonucleotides having a desirable bias of random codon sequences that are operationally linked to expression sequences. The 15 invention provides for methods of constructing such populations of cells as well.

Random oligonucleotides synthesized by any of the methods described previously can also be expressed on the surface of filamentous bacteriophage, such as M13, for 20 example, without the joining together of precursor oligonucleotides. A vector such as that shown in Figure 4, M13IX30, can be used. This vector exhibits all the functional features of the combined vector shown in Figure 3C for surface expression of gVIII-peptide fusion proteins. 25 The complete nucleotide sequence for M13IX30 (SEQ ID NO: 3) is shown in Figure 7.

M13IX30 contains a wild type gVIII for phage viability and a pseudo gVIII sequence for peptide fusions. The vector also contains in frame restriction sites for cloning 30 random peptides. The cloning sites in this vector are Xho I, Stu I and Spe I. Oligonucleotides should therefore be synthesized with the appropriate complementary ends for annealing and ligation or insertional mutagenesis. Alternatively, the appropriate termini can be generated by

PCR technology. Between the restriction sites and the pseudo gVIII sequence is an in-frame amber stop codon, again, ensuring complete viability of phage in constructing and manipulating the library. Expression and screening is 5 performed as described above for the surface expression library of oligonucleotides generated from precursor portions.

Thus, the invention provides a method of selecting peptides capable of being bound by a ligand binding protein 10 from a population of random peptides by (a) operationally linking a diverse population of oligonucleotides having a desirable bias of random codon sequences to expression elements; (b) introducing said population of vectors into a compatible host under conditions sufficient for 15 expressing said population of random peptides; and (c) determining the peptides which bind to said binding protein. Also provided is a method for determining the encoding nucleic acid sequence of such selected peptides.

The following examples are intended to illustrate, but 20 not limit the invention.

EXAMPLE I

Isolation and Characterization of Peptide Ligands Generated From Right and Left Half Random Oligonucleotides

25 This example shows the synthesis of random oligonucleotides and the construction and expression of surface expression libraries of the encoded randomized peptides. The random peptides of this example derive from the mixing and joining together of two random 30 oligonucleotides. Also demonstrated is the isolation and characterization of peptide ligands and their corresponding nucleotide sequence for specific binding proteins.

Synthesis of Random Oligonucleotides

The synthesis of two randomized oligonucleotides which correspond to smaller portions of a larger randomized oligonucleotide is shown below. Each of the two smaller portions make up one-half of the larger oligonucleotide. The population of randomized oligonucleotides constituting each half are designated the right and left half. Each population of right and left halves are ten codons in length with twenty random codons at each position. The right half corresponds to the sense sequence of the randomized oligonucleotides and encode the carboxy terminal half of the expressed peptides. The left half corresponds to the anti-sense sequence of the randomized oligonucleotides and encode the amino terminal half of the expressed peptides. The right and left halves of the randomized oligonucleotide populations are cloned into separate vector species and then mixed and joined so that the right and left halves core together in random combination to produce a single expression vector species which contains a population of randomized oligonucleotides twenty codons in length. Electroporation of the vector population into an appropriate host produces filamentous phage which express the random peptides on their surface.

The reaction vessels for oligonucleotide synthesis were obtained from the manufacturer of the automated synthesizer (Millipore, Burlington, MA; supplier of MilliGen/Bioscience Cyclone Plus Synthesizer). The vessels were supplied as packages containing empty reaction columns (1 μ mole), frits, crimps and plugs (MilliGen/Bioscience catalog # GEN 860458). Derivatized and underderivatized control pore glass, phosphoramidite nucleotides, and synthesis reagents were also obtained from MilliGen/Bioscience. Crimper and decrimper tools were obtained from Fisher Scientific Co., Pittsburgh, PA (Catalog numbers 06-406-20 and 06-406-25A, respectively).

Ten reaction columns were used for right half synthesis of random oligonucleotides ten codons in length. The oligonucleotides have 5 monomers at their 3' end of the sequence 5'GAGCT3' and 6 monomers at their 5' end of the sequence 5'AATTCCAT3'. The synthesizer was fitted with a column derivatized with a thymine nucleotide (T-column, MilliGen/Bioscience # 0615.50) and was programmed to synthesize the sequences shown in Table I for each of ten columns in independent reaction sets. The sequence of the last three monomers (from right to left since synthesis proceeds 3' to 5') encode the indicated amino acids:

Table I

	<u>Column</u>	<u>Sequence (5' to 3')</u>	<u>Amino Acids</u>
15	column 1R	(T/G)TTGAGCT	Phe and Val
	column 2R	(T/C)CTGAGCT	Ser and Pro
	column 3R	(T/C)ATGAGCT	Tyr and His
	column 4R	(T/C)GTGAGCT	Cys and Arg
	column 5R	(C/A)TGGAGCT	Leu and Met
20	column 6R	(C/G)AGGAGCT	Gln and Glu
	column 7R	(A/G)CTGAGCT	Thr and Ala
	column 8R	(A/G)ATGAGCT	Asn and Asp
	column 9R	(T/G)GGGAGCT	Trp and Gly
	column 1R	A(T/A)AGAGCT	Ile and Cys

25 where the two monomers in parentheses denote a single monomer position within the codon and indicate that an equal mixture of each monomer was added to the reaction for coupling. The monomer coupling reactions for each of the 10 columns were performed as recommended by the 30 manufacturer (amidite version S1.06, # 8400-050990, scale 1 μ M). After the last coupling reaction, the columns were washed with acetonitrile and lyophilized to dryness.

Following synthesis, the plugs were removed from each

column using a decriper and the reaction products were poured into a single weigh boat. Initially the bead mass increases, due to the weight of the monomers, however, at later rounds of synthesis material is lost. In either 5 case, the material was equalized with underivatized control pore glass and mixed thoroughly to obtain a random distribution of all twenty codon species. The reaction products were then aliquotted into 10 new reaction columns by removing 25 mg of material at a time and placing it into 10 separate reaction columns. Alternatively, the reaction products can be aliquotted by suspending the beads in a liquid that is dense enough for the beads to remain dispersed, preferably a liquid that is equal in density to the beads, and then aliquoting equal volumes of the 15 suspension into separate reaction columns. The lip on the inside of the columns where the frits rest was cleared of material using vacuum suction with a syringe and 25 G needle. New frits were placed onto the lips, the plugs were fitted into the columns and were crimped into place 20 using a crimper.

Synthesis of the second codon position was achieved using the above 10 columns containing the random mixture of reaction products from the first codon synthesis. The monomer coupling reactions for the second codon position 25 are shown in Table II. An A in the first position means that any monomer can be programmed into the synthesizer. At that position, the first monomer position is not coupled by the synthesizer since the software assumes that the monomer is already attached to the column. An A also 30 denotes that the columns from the previous codon synthesis should be placed on the synthesizer for use in the present synthesis round. Reactions were again sequentially repeated for each column as shown in Table II and the reaction products washed and dried as described above.

Table II

<u>Column</u>	<u>Sequence (5' to 3')</u>	<u>Amino Acids</u>
5	column 1R (T/G)TTA	Phe and Val
	column 2R (T/C)CTA	Ser and Pro
	column 3R (T/C)ATA	Tyr and His
	column 4R (T/C)GTA	Cys and Arg
	column 5R (C/A)TGA	Leu and Met
10	column 6R (C/G)AGA	Gln and Glu
	column 7R (A/G)CTA	Thr and Ala
	column 8R (A/G)ATA	Asn and Asp
	column 9R (T/G)GGA	Trp and Gly
	column 10R A(T/A)AA	Ile and Cys

Randomization of the second codon position was achieved by
 15 removing the reaction products from each of the columns and
 thoroughly mixing the material. The material was again
 divided into new reaction columns and prepared for monomer
 coupling reactions as described above.

Random synthesis of the next seven codons (positions
 20 3 through 9) proceeded identically to the cycle described
 above for the second codon position and again used the
 monomer sequences of Table II. Each of the newly repacked
 columns containing the random mixture of reaction products
 from synthesis of the previous codon position was used for
 25 the synthesis of the subsequent codon position. After
 synthesis of the codon at position nine and mixing of the
 reaction products, the material was divided and repacked
 into 40 different columns and the monomer sequences shown
 in Table III were coupled to each of the 40 columns in
 30 independent reactions. The oligonucleotides from each of
 the 40 columns were mixed once more and cleaved from the
 control pore glass as recommended by the manufacturer.

Table III

	<u>Column</u>	<u>Sequence (5' to 3')</u>
5	column 1R	AATTCTTTA
	column 2R	AATTCTGTTA
	column 3R	AATTCGTTA
	column 4R	AATTCCGGTTA
	column 5R	AATTCTTCTA
	column 6R	AATTCTCCTA
10	column 7R	AATTCGTCTA
	column 8R	AATTCGCCTA
	column 9R	AATTCTTATA
	column 10R	AATTCTCATA
	column 11R	AATTCGTATA
	column 12R	AATTCGCATA
15	column 13R	AATTCTTGTA
	column 14R	AATTCTCGTA
	column 15R	AATTCGTGT
	column 16R	AATTCGCGTA
	column 17R	AATTCTCTGA
	column 18R	AATTCTATGA
20	column 19R	AATTCGCTGA
	column 20R	AATTCGATGA
	column 21R	AATTCTCAGA
	column 22R	AATTCTGAGA
	column 23R	AATTCGCAGA
	column 24R	AATTCGGAGA
25	column 25R	AATTCTACTA
	column 26R	AATTCTGCTA
	column 27R	AATTCGACTA
	column 28R	AATTCGGCTA
	column 29R	AATTCTAATA
	column 30R	AATTCTGATA
30	column 31R	AATTCGAATA
	column 32R	AATTCGGATA
	column 33R	AATTCTTGGA

	column 34R	AATTCTGGGA
	column 35R	AATT C GTGG <u>A</u>
	column 36R	AATT C GGGG <u>A</u>
	column 37R	AATTCTATA <u>A</u>
5	column 38R	AATTCTAAAA
	column 39R	AATT C GATA <u>A</u>
	column 40R	AATT C GAAAA <u>A</u>

Left half synthesis of random oligonucleotides proceeded similarly to the right half synthesis. This half 10 of the oligonucleotide corresponds to the anti-sense sequence of the encoded randomized peptides. Thus, the complementary sequence of the codons in Tables I through III are synthesized. The left half oligonucleotides also have 5 monomers at their 3' end of the sequence 5'GAGCT3' 15 and 8 monomers at their 5' end of the sequence 5'AATTCCAT3'. The rounds of synthesis, washing, drying, mixing, and dividing are as described above.

For the first codon position, the synthesizer was fitted with a T-column and programmed to synthesize the 20 sequences shown in Table IV for each of ten columns in independent reaction sets. As with right half synthesis, the sequence of the last three monomers (from right to left) encode the indicated amino acids:

Table IV

	<u>Column</u>	<u>Sequence (5' to 3')</u>	<u>Amino Acids</u>
5	column 1L	AA(A/C)GAGCT	Phe and Val
	column 2L	AG(A/G)GAGCT	Ser and Pro
	column 3L	AT(A/G)GACCT	Tyr and His
	column 4L	AC(A/G)GAGCT	Cys and Arg
	column 5L	CA(G/T)GAGCT	Leu and Met
	column 6L	CT(G/C)GAGCT	Gln and Glu
	column 7L	AG(T/C)GAGCT	Thr and Ala
	column 8L	AT(T/C)GAGCT	Asn and Asp
	column 9L	CC(A/C)GAGCT	Trp and Gly
	column 10L	T(A/T)TGAGCT	Ile and Cys

Following washing and drying, the plugs for each column
15 were removed, mixed and aliquotted into ten new reaction
columns as described above. Synthesis of the second codon
position was achieved using these ten columns containing
the random mixture of reaction products from the first
codon synthesis. The monomer coupling reactions for the
20 second codon position are shown in Table V.

Table V

	<u>Column</u>	<u>Sequence (5' to 3')</u>	<u>Amino Acids</u>
25	column 1L	AA(A/C)A	Phe and Val
	column 2L	AG(A/G)A	Ser and Pro
	column 3L	AT(A/G)A	Tyr and His
	column 4L	AC(A/G)A	Cys and Arg
	column 5L	CA(G/T)A	Leu and Met
	column 6L	CT(G/C)A	Gln and Glu
	column 7L	AG(T/C)A	Thr and Ala
	column 8L	AT(T/C)A	Asn and Asp
	column 9L	CC(A/C)A	Trp and Gly
	column 10L	T(A/T)TA	Ile and Cys

Again, randomization of the second codon position was achieved by removing the reaction products from each of the columns and thoroughly mixing the beads. The beads were repacked into ten new reaction columns.

5 Random synthesis of the next seven codon positions proceeded identically to the cycle described above for the second codon position and again used the monomer sequences of Table V. After synthesis of the codon at position nine and mixing of the reaction products, the material was 10 divided and repacked into 40 different columns and the monomer sequences shown in Table VI were coupled to each of the 40 columns in independent reactions.

Table VI

	<u>Column</u>	<u>Sequence (5' to 3')</u>
15	column 1L	AATTCCATAAA XXA
	column 2L	AATTCCATAAAC XXA
	column 3L	AATTCCATAAC XXA
	column 4L	AATTCCATAACC XXA
20	column 5L	AATTCCATAGA XXA
	column 6L	AATTCCATAGAC XXA
	column 7L	AATTCCATAGG XXA
	column 8L	AATTCCATAGGC XXA
	column 9L	AATTCCATATA XXA
25	column 10L	AATTCCATATA XXA
	column 11L	AATTCCATATG XXA
	column 12L	AATTCCATATGC XXA
	column 13L	AATTCCATACA XXA
	column 14L	AATTCCATACAC XXA
30	column 15L	AATTCCATACG XXA
	column 16L	AATTCCATACGC XXA
	column 17L	AATTCCATCAG XXA
	column 18L	AATTCCATCAGC XXA
	column 19L	AATTCCATCAT XXA
35	column 20L	AATTCCATCATC XXA

	column 21L	AATTCCATCTGAXXA
	column 22L	AATTCCATCTGCXXA
	column 23L	AATTCCATCTCAXXA
	column 24L	AATTCCATCTCCXXA
5	column 25L	AATTCCATAGTAXXA
	column 26L	AATTCCATAGTCXXA
	column 27L	AATTCCATAGCAXXA
	column 28L	AATTCCATAGCCXXA
	column 29L	AATTCCATATTAXXA
10	column 30L	AATTCCATATTCCXXA
	column 31L	AATTCCATATCAXXA
	column 32L	AATTCCATATCCXXA
	column 33L	AATTCCATCCAAXXA
	column 34L	AATTCCATCCACXXA
15	column 35L	AATTCCATCCCAXXA
	column 36L	AATTCCATCCCCXXA
	column 37L	AATTCCATTATAAXXA
	column 38L	AATTCCATTATCXXA
	column 39L	AATTCCATTTTAXXA
20	column 40L	AATTCCATTTTCXXA

The first two monomers denoted by an "X" represent an equal mixture of all four nucleotides at that position. This is necessary to retain a relatively unbiased codon sequence at the junction between right and left half oligonucleotides.

25 The above right and left half random oligonucleotides were cleaved and purified from the supports and used in constructing the surface expression libraries below.

Vector Construction

Two M13-based vectors, M13IX42 (SEQ ID NO: 1) and 30 M13IX22 (SEQ ID NO: 2), were constructed for the cloning and propagation of right and left half populations of random oligonucleotides, respectively. The vectors were specially constructed to facilitate the random joining and subsequent expression of right and left half

oligonucleotide populations. Each vector within the population contains one right and one left half oligonucleotide from the population joined together to form a single contiguous oligonucleotide with random codons which is twenty-two codons in length. The resultant population of vectors are used to construct a surface expression library.

M13IX42, or the right-half vector, was constructed to harbor the right half populations of randomized oligonucleotides. M13mp18 (Pharmacia, Piscataway NJ) was the starting vector. This vector was genetically modified to contain, in addition to the encoded wild type M13 gene VIII already present in the vector: (1) a pseudo-wild type M13 gene VIII sequence with a stop codon (amber) placed between it and an Eco RI-Sac I cloning site for randomized oligonucleotides; (2) a pair of Fok I sites to be used for joining with M13IX22, the left-half vector; (3) a second amber stop codon placed on the opposite side of the vector than the portion being combined with the left-half vector; and (4) various other mutations to remove redundant restriction sites and the amino terminal portion of Lac Z.

The pseudo-wild type M13 gene VIII was used for surface expression of random peptides. The pseudo-wild type gene encodes the identical amino acid sequence as that of the wild type gene; however, the nucleotide sequence has been altered so that only 63% identity exists between this gene and the encoded wild type gene VIII. Modification of the gene VIII nucleotide sequence used for surface expression reduces the possibility of homologous recombination with the wild type gene VIII contained on the same vector. Additionally, the wild type M13 gene VIII was retained in the vector system to ensure that at least some functional, non-fusion coat protein would be produced. The inclusion of wild type gene VIII therefore reduces the possibility of non-viable phage production from the random

peptide fusion genes.

The pseudo-wild type gene VIII was constructed by chemically synthesizing a series of oligonucleotides which encode both strands of the gene. The oligonucleotides are 5 presented in Table VII (SEQ ID NOS: 7 through 16).

TABLE VII

Pseudo-Wild Type Gene VIII Oligonucleotide Series

	<u>Top Strand Oligonucleotides</u>	<u>Sequence (5' to 3')</u>
10	VIII 03	GATCC TAG GCT GAA GGC GAT GAC CCT GCT AAG GCT GC
	VIII 04	A TTC AAT AGT TTA CAG GCA AGT GCT ACT GAG TAC A
	VIII 05	TT GGC TAC GCT TGG GCT ATG GTA GTA GTT ATA GTT
15	VIII 06	GGT GCT ACC ATA GGG ATT AAA TTA TTC AAA AAG TT
	VIII 07	T ACG AGC AAG GCT TCT TA
20	<u>Bottom Strand Oligonucleotides</u>	
	VIII 08	AGC TTA AGA AGC CTT GCT CGT AAA CTT TTT GAA TAA TTT
	VIII 09	AAT CCC TAT GGT AGC ACC AAC TAT AAC TAC TAC CAT
25	VIII 10	AGC CCA AGC GTA GCC AAT GTA CTC AGT AGC ACT TG
	VIII 11	C CTG TAA ACT ATT GAA TGC AGC CTT AGC AGG GTC
	VIII 12	ATC GCC TTC AGC CTA G

30 Except for the terminal oligonucleotides VIII 03 (SEQ

ID NO: 7) and VIII 08 (SEQ ID NO: 12), the above oligonucleotides (oligonucleotides VIII 04-VIII 07 and 09-12 (SEQ ID NOS: 8 through 11 and 13 through 16)) were mixed at 200 ng each in 10 μ l final volume and phosphorylated with T4 polynucleotide Kinase (Pharmacia, Piscataway, NJ) with 1 mM ATP at 37°C for 1 hour. The reaction was stopped at 65°C for 5 minutes. Terminal oligonucleotides were added to the mixture and annealed into double-stranded form by heating to 65°C for 5 minutes, followed by cooling to room temperature over a period of 30 minutes. The annealed oligonucleotides were ligated together with 1.0 U of T4 DNA ligase (BRL). The annealed and ligated oligonucleotides yield a double-stranded DNA flanked by a Bam HI site at its 5' end and by a Hind III site at its 3' end. A translational stop codon (amber) immediately follows the Bam HI site. The gene VIII sequence begins with the codon GAA (Glu) two codons 3' to the stop codon. The double-stranded insert was phosphorylated using T4 DNA Kinase (Pharmacia, Piscataway, NJ) and ATP (10 mM Tris-HCl, pH 7.5, 10 mM MgCl₂) and cloned in frame with the Eco RI and Sac I sites within the M13 polylinker. To do so, M13mp18 was digested with Bam HI (New England Biolabs, Beverley, MA) and Hind III (New England Biolabs) and combined at a molar ratio of 1:10 with the double-stranded insert. The ligations were performed at 16°C overnight in 1X ligase buffer (50 mM Tris-HCl, pH 7.8, 10 mM MgCl₂, 20 mM DTT, 1 mM ATP, 50 μ g/ml BSA) containing 1.0 U of T4 DNA ligase (New England Biolabs). The ligation mixture was transformed into a host and screened for positive clones using standard procedures in the art.

Several mutations were generated within the right-half vector to yield functional M13IX42. The mutations were generated using the method of Kunkel et al., *Meth. Enzymol.* 154:367-382 (1987), which is incorporated herein by reference, for site-directed mutagenesis. The reagents, strains and protocols were obtained from a Bio Rad

Mutagenesis kit (Bio Rad, Richmond, CA) and mutagenesis was performed as recommended by the manufacturer.

A Fok I site used for joining the right and left halves was generated 8 nucleotides 5' to the unique Eco RI site using the oligonucleotide 5'-CTCGAATTCTGACATCCTGGTCATAGC-3' (SEQ ID NO: 17). The second Fok I site retained in the vector is naturally encoded at position 3547; however, the sequence within the overhang was changed to encode CTTC. Two Fok I sites were removed from the vector at positions 239 and 7244 of M13mp18 as well as the Hind III site at the end of the pseudo gene VIII sequence using the mutant oligonucleotides 5'-CATTTCAGATGGCTAGA-3' (SEQ ID NO: 18) and 5'-TAGCATTAAACGTCCAATA-3' (SEQ ID NO: 19), respectively. New Hind III and Mlu I sites were also introduced at position 3919 and 3951 of M13IX42. The oligonucleotides used for this mutagenesis had the sequences 5'-ATATATTTAGTAAGCTTCATCTTCT-3' (SEQ ID NO: 20) and 5'-GACAAAGAACGCGTGAAACTTT-3' (SEQ ID NO: 21), respectively. The amino terminal portion of Lac Z was deleted by oligonucleotide-directed mutagenesis using the mutant oligonucleotide 5'-GCAGGCCTCTCGCTATTGCTTAAGAAGCCTTGCT-3' (SEQ ID NO: 22). This deletion also removed a third M13mp18 derived Fok I site. The distance between the Eco RI and Sac I sites was increased to ensure complete double digestion by inserting a spacer sequence. The spacer sequence was inserted using the oligonucleotide 5'-TTCAGCCTAGGATCCGCCGAGCTCTCCTACCTGCGAATTCTGACATCC-3' (SEQ ID NO: 23). Finally, an amber stop codon was placed at position 4492 using the mutant oligonucleotide 5'-TGGATTATACTTCTAAATAATGGA-3' (SEQ ID NO: 24). The amber stop codon is used as a biological selection to ensure the proper recombination of vector sequences to bring together right and left halves of the randomized oligonucleotides. In constructing the above mutations, all changes made in a M13 coding region were performed such that the amino acid

sequence remained unaltered. It should be noted that several mutations within M13mp18 were found which differed from the published sequence. Where known, these sequence differences are recorded herein as found and therefore may 5 not correspond exactly to the published sequence of M13mp18.

The sequence of the resultant vector, M13IX42, is shown in Figure 5 (SEQ ID NO: 1). Figure 3A also shows M13IX42 where each of the elements necessary for producing 10 a surface expression library between right and left half randomized oligonucleotides is marked. The sequence between the two Fok I sites shown by the arrow is the portion of M13IX42 which is to be combined with a portion of the left-half vector to produce random oligonucleotides 15 as fusion proteins of gene VIII.

M13IX22, or the left-half vector, was constructed to harbor the left half populations of randomized oligonucleotides. This vector was constructed from M13mp19 (Pharmacia, Piscataway, NJ) and contains: (1) Two Fok I 20 sites for mixing with M13IX42 to bring together the left and right halves of the randomized oligonucleotides; (2) sequences necessary for expression such as a promoter and signal sequence and translation initiation signals; (3) an Eco RI-Sac I cloning site for the randomized 25 oligonucleotides; and (4) an amber stop codon for biological selection in bringing together right and left half oligonucleotides.

Of the two Fok I sites used for mixing M13IX22 with M13IX42, one is naturally encoded in M13mp18 and M13mp19 30 (at position 3547). As with M13IX42, the overhang within this naturally occurring Fok I site was changed to CTTC. The other Fok I site was introduced after construction of the translation initiation signals by site-directed mutagenesis using the oligonucleotide 5'-

TAACACTCATTCCGGATGGAA-TCTGGAGTCTGGGT-3' (SEQ ID NO: 25).

The translation initiation signals were constructed by annealing of overlapping oligonucleotides as described above to produce a double-stranded insert containing a 5' 5 Eco RI site and a 3' Hind III site. The overlapping oligonucleotides are shown in Table VIII (SEQ ID NOS: 26 through 34) and were ligated as a double-stranded insert between the Eco RI and Hind III sites of M13mp18 as described for the pseudo gene VIII insert. The ribosome 10 binding site (AGGAGAC) is located in oligonucleotide 015 (SEQ ID NO: 26) and the translation initiation codon (ATG) is the first three nucleotides of oligonucleotide 016 (SEQ ID NO: 27).

TABLE VIII

15 oligonucleotide Series for construction of
Translation Signals in M13IX22

	<u>oligonucleotide</u>	<u>Sequence (5' to 3')</u>
	015	AATT C GCC AAG GAG ACA GTC AT
	016	AATG AAA TAC CTA TTG CCT ACG GCA
20	017	GCC GCT GGA TTG TT
	018	ATTA CTC GCT GCC CAA CCA GCC ATG
	019	GCC GAG CTC GTG AT
25	020	GACC CAG ACT CCA GATATC CAA CAG
	021	GAA TGA GTG TTA AT
	022	TCT AGA ACG CGT C
30	023	ACGT G ACG CGT TCT AGA AT TAA
		CACTCA TTC CTG T
		TG GAT ATC TGG AGT CTG GGT CAT
		CAC GAG CTC GGC CAT G
		GC TGG TTG GGC AGC GAG TAA TAA
		CAA TCC AGC GGC TGC C
		GT AGG CAA TAG GTA TTT CAT TAT
		GAC TGT CCT TGG CG

oligonucleotide 017 (SEQ ID NO: 27) contained a Sac I restriction site 67 nucleotides downstream from the ATG codon. The naturally occurring Eco RI site was removed and a new site introduced 25 nucleotides downstream from the 5 Sac I. Oligonucleotides 5'-TGACTGTCTCCTGGCGTGTGAAATTGTTA-
3' (SEQ ID NO: 35) and 5'-TAACACTCATTCCGGATGGAATTCTGGAGTCT
GGGT-3' (SEQ ID NO: 36) were used to generate each of the 10 mutations, respectively. An amber stop codon was also introduced at position 3263 of M13mp18 using the oligonucleotide 5'-CAATTCTATCCTAAATCTTACCAAC-3' (SEQ ID NO: 15 37).

In addition to the above mutations, a variety of other modifications were made to remove certain sequences and redundant restriction sites. The LAC Z ribosome binding 15 site was removed when the original Eco RI site in M13mp18 was mutated. Also, the Fok I sites at positions 239, 6361 and 7244 of M13mp18 were likewise removed with mutant oligonucleotides 5'-CATTTCAGATGGCTTAGA-3' (SEQ ID NO: 20 38), 5'-CGAAAGGGGGTGTGCTGCAA-3' (SEQ ID NO: 39) and 5'-TAGCATTAAACGTCCAATA-3' (SEQ ID NO: 40), respectively. Again, mutations within the coding region did not alter the amino acid sequence.

The resultant vector, M13IX22, is 7320 base pairs in 25 length, the sequence of which is shown in Figure 6 (SEQ ID NO: 2). The Sac I and Eco RI cloning sites are at positions 6290 and 6314, respectively. Figure 3A also shows M13IX22 where each of the elements necessary for producing a surface expression library between right and left half randomized oligonucleotides is marked.

30 Library Construction

Each population of right and left half randomized oligonucleotides from columns 1R through 40R and columns 1L through 40L are cloned separately into M13IX42 and M13IX22,

respectively, to create sublibraries of right and left half randomized oligonucleotides. Therefore, a total of eighty sublibraries are generated. Separately maintaining each population of randomized oligonucleotides until the final screening step is performed to ensure maximum efficiency of annealing of right and left half oligonucleotides. The greater efficiency increases the total number of randomized oligonucleotides which can be obtained. Alternatively, one can combine all forty populations of right half oligonucleotides (columns 1R-40R) into one population and of left half oligonucleotides (columns 1L-40L) into a second population to generate just one sublibrary for each.

For the generation of sublibraries, each of the above populations of randomized oligonucleotides are cloned separately into the appropriate vector. The right half oligonucleotides are cloned into M13IX42 to generate sublibraries M13IX42.1R through M13IX42.40R. The left half oligonucleotides are similarly cloned into M13IX22 to generate sublibraries M13IX22.1L through M13IX22.40L. Each vector contains unique Eco RI and Sac I restriction enzyme sites which produce 5' and 3' single-stranded overhangs, respectively, when digested. The single strand overhangs are used for the annealing and ligation of the complementary single-stranded random oligonucleotides.

The randomized oligonucleotide populations are cloned between the Eco RI and Sac I sites by sequential digestion and ligation steps. Each vector is treated with an excess of Eco RI (New England Biolabs) at 37°C for 2 hours followed by addition of 4-24 units of calf intestinal alkaline phosphatase (Boehringer Mannheim, Indianapolis, IN). Reactions are stopped by phenol/chloroform extraction and ethanol precipitation. The pellets are resuspended in an appropriate amount of distilled or deionized water (dH₂O). About 10 pmol of vector is mixed with a 5000-fold molar excess of each population of randomized

oligonucleotides in 10 μ l of 1X ligase buffer (50 mM Tris-HCl, pH 7.8, 10 mM MgCl₂, 20 mM DTT, 1 mM ATP, 50 μ g/ml BSA) containing 1.0 U of T4 DNA ligase (BRL, Gaithersburg, MD). The ligation is incubated at 16°C for 16 hours. Reactions 5 are stopped by heating at 75°C for 15 minutes and the DNA is digested with an excess of Sac I (New England Biolabs) for 2 hours. Sac I is inactivated by heating at 75°C for 15 minutes and the volume of the reaction mixture is adjusted to 300 μ l with an appropriate amount of 10X ligase 10 buffer and dH₂O. One unit of T4 DNA ligase (BRL) is added and the mixture is incubated overnight at 16°C. The DNA is ethanol precipitated and resuspended in TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). DNA from each ligation is electroporated into XL1 Blue™ cells (Stratagene, La Jolla, 15 CA), as described below, to generate the sublibraries.

E. coli XL1 Blue™ is electroporated as described by Smith et al., Focus 12:38-40 (1990) which is incorporated herein by reference. The cells are prepared by inoculating a fresh colony of XL1s into 5 mls of SOB without magnesium 20 (20 g bacto-tryptone, 5 g bacto-yeast extract, 0.584 g NaCl, 0.186 g KCl, dH₂O to 1,000 mls) and grown with vigorous aeration overnight at 37°C. SOB without magnesium (500 ml) is inoculated at 1:1000 with the overnight culture 25 and grown with vigorous aeration at 37°C until the OD₅₅₀ is 0.8 (about 2 to 3 h). The cells are harvested by centrifugation at 5,000 rpm (2,600 x g) in a GS3 rotor (Sorvall, Newtown, CT) at 4°C for 10 minutes, resuspended in 500 ml of ice-cold 10% (v/v) sterile glycerol and 30 centrifuged and resuspended a second time in the same manner. After a third centrifugation, the cells are resuspended in 10% sterile glycerol at a final volume of about 2 ml, such that the OD₅₅₀ of the suspension is 200 to 300. Usually, resuspension is achieved in the 10% glycerol 35 that remains in the bottle after pouring off the supernate. Cells are frozen in 40 μ l aliquots in microcentrifuge tubes using a dry ice-ethanol bath and stored frozen at -70°C.

Frozen cells are electroporated by thawing slowly on ice before use and mixing with about 10 pg to 500 ng of vector per 40 μ l of cell suspension. A 40 μ l aliquot is placed in an 0.1 cm electroporation chamber (Bio-Rad, 5 Richmond, CA) and pulsed once at 0°C using 200 Ω parallel resistor, 25 μ F, 1.88 kV, which gives a pulse length (τ) of ~4 ms. A 10 μ l aliquot of the pulsed cells are diluted into 1 ml SOC (98 mls SOB plus 1 ml of 2 M MgCl₂ and 1 ml of 10 2 M glucose) in a 12- x 75-mm culture tube, and the culture is shaken at 37°C for 1 hour prior to culturing in selective media, (see below).

Each of the eighty sublibraries are cultured using methods known to one skilled in the art. Such methods can be found in Sanbrook et al., Molecular Cloning: A 15 Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, 1989, and in Ausubel et al., Current Protocols in Molecular Biology, John Wiley and Sons, New York, 1989, both of which are incorporated herein by reference. Briefly, the above 1 ml sublibrary cultures 20 were grown up by diluting 50-fold into 2XYT media (16 g tryptone, 10 g yeast extract, 5 g NaCl) and culturing at 37°C for 5-8 hours. The bacteria were pelleted by centrifugation at 10,000 xg. The supernatant containing phage was transferred to a sterile tube and stored at 4°C.

25 Double strand vector DNA containing right and left half randomized oligonucleotide inserts is isolated from the cell pellet of each sublibrary. Briefly, the pellet is washed in TE (10 mM Tris, pH 8.0, 1 mM EDTA) and 30 recollected by centrifugation at 7,000 rpm for 5' in a Sorval centrifuge (Newtown, CT). Pellets are resuspended in 6 mls of 10% Sucrose, 50 mM Tris, pH 8.0. 3.0 ml of 10 mg/ μ l lysozyme is added and incubated on ice for 20 minutes. 12 mls of 0.2 M NaOH, 1% SDS is added followed by 35 10 minutes on ice. The suspensions are then incubated on ice for 20 minutes after addition of 7.5 mls of 3 M NaOAc,

pH 4.6. The samples are centrifuged at 15,000 rpm for 15 minutes at 4°C, RNased and extracted with phenol/chloroform, followed by ethanol precipitation. The pellets are resuspended, weighed and an equal weight of 5 CsCl_2 is dissolved into each tube until a density of 1.60 g/ml is achieved. EtBr is added to 600 $\mu\text{g}/\text{ml}$ and the double-stranded DNA is isolated by equilibrium centrifugation in a TV-1665 rotor (Sorval) at 50,000 rpm for 6 hours. These DNAs from each right and left half 10 sublibrary are used to generate forty libraries in which the right and left halves of the randomized oligonucleotides have been randomly joined together.

Each of the forty libraries are produced by joining together one right half and one left half sublibrary. The 15 two sublibraries joined together corresponded to the same column number for right and left half random oligonucleotide synthesis. For example, sublibrary M13IX42.1R is joined with M13IX22.1L to produce the surface expression library M13IX.1RL. In the alternative situation 20 where only two sublibraries are generated from the combined populations of all right half synthesis and all left half synthesis, only one surface expression library would be produced.

For the random joining of each right and left half 25 oligonucleotide populations into a single surface expression vector species, the DNAs isolated from each sublibrary are digested an excess of Fok I (New England Biolabs). The reactions are stopped by phenol/chloroform extraction, followed by ethanol precipitation. Pellets are 30 resuspended in dH_2O . Each surface expression library is generated by ligating equal molar amounts (5-10 pmol) of Fok I digested DNA isolated from corresponding right and left half sublibraries in 10 μl of 1X ligase buffer containing 1.0 U of T4 DNA ligase (Bethesda Research 35 Laboratories, Gaithersburg, MD). The ligations proceed

overnight at 16°C and are electroporated into the sup 0 strain MK30-3 (Boehringer Mannheim Biochemical, (BMB), Indianapolis, IN) as previously described for XL1 cells. Because MK30-3 is sup 0, only the vector portions encoding the randomized oligonucleotides which come together will produce viable phage.

Screening of Surface Expression Libraries

Purified phage are prepared from 50 ml liquid cultures of XL1 Blue™ cells (Stratagene) which are infected at a m.o.i. of 10 from the phage stocks stored at 4°C. The cultures are induced with 2 mM IPTG. Supernatants from all cultures are combined and cleared by two centrifugations, and the phage are precipitated by adding 1/7.5 volumes of PEG solution (25% PEG-8000, 2.5 M NaCl), followed by incubation at 4°C overnight. The precipitate is recovered by centrifugation for 90 minutes at 10,000 x g. Phage pellets are resuspended in ~5 ml of 0.01 M Tris-HCl, pH 7.6, 1.0 mM EDTA, and 0.1% Sarkosyl and then shaken slowly at room temperature for 30 minutes. The solutions are adjusted to 0.5 M NaCl and to a final concentration of 5% polyethylene glycol. After 2 hours at 4°C, the precipitates containing the phage are recovered by centrifugation for 1 hour at 15,000 X g. The precipitates are resuspended in 10 ml of NET buffer (0.1 M NaCl, 1.0 mM EDTA, and 0.01 M Tris-HCl, pH 7.6), mixed well, and the phage repelleted by centrifugation at 170,000 X g for 3 hours. The phage pellets are subsequently resuspended overnight in 2 ml of NET buffer and subjected to cesium chloride centrifugation for 18 hours at 110,000 X g (3.86 g of cesium chloride in 10 ml of buffer). Phage bands are collected, diluted 7-fold with NET buffer, re-centrifuged at 170,000 X g for 3 hours, resuspended, and stored at 4°C in 0.3 ml of NET buffer containing 0.1 mM sodium azide.

Ligand binding proteins used for panning on

streptavidin coated dishes are first biotinylated and then absorbed against UV-inactivated blocking phage (see below). The biotinylating reagents are dissolved in dimethylformamide at a ratio of 2.4 mg solid NHS-SS-Biotin (sulfosuccinimidyl 2-(biotinamido)ethyl-1,3'-dithiopropionate; Pierce, Rockford, IL) to 1 ml solvent and used as recommended by the manufacturer. Small-scale reactions are accomplished by mixing 1 μ l dissolved reagent with 43 μ l of 1 mg/ml ligand binding protein diluted in 10 sterile bicarbonate buffer (0.1 M NaHCO₃, pH 8.6). After 2 hours at 25°C, residual biotinylating reagent is reacted with 500 μ l 1 M ethanolamine (pH adjusted to 9 with HCl) for an additional 2 hours. The entire sample is diluted with 1 ml TBS containing 1 mg/ml BSA, concentrated to about 15 50 μ l on a Centricon 30 ultra-filter (Amicon), and washed on the same filter three times with 2 ml TBS and once with 1 ml TBS containing 0.02% NaN₃ and 7 \times 10¹² UV-inactivated blocking phage (see below); the final retentate (60-80 μ l) is stored at 4°C. Ligand binding proteins biotinylated 20 with the NHS-SS-Biotin reagent are linked to biotin via a disulfide-containing chain.

UV-irradiated M13 phage were used for blocking binding proteins which fortuitously bound filamentous phage in general. M13mp8 (Messing and Vieira, Gene 19: 262-276, 25 (1982), which is incorporated herein by reference) was chosen because it carries two amber stop codons, which ensure that the few phage surviving irradiation will not grow in the sup O strains used to titer the surface expression libraries. A 5 ml sample containing 5 \times 10¹³ 30 M13mp8 phage, purified as described above, was placed in a small petri plate and irradiated with a germicidal lamp at a distance of two feet for 7 minutes (flux 150 μ W/cm²). NaN₃ was added to 0.02% and phage particles concentrated to 10¹⁴ particles/ml on a Centricon 30-kDa ultrafilter 35 (Amicon).

For panning, polystyrene petri plates (60 x 15 mm, Falcon; Becton Dickinson, Lincoln Park, NJ) are incubated with 1 ml of 1 mg/ml of streptavidin (BMB) in 0.1 M NaHCO₃ pH 8.6-0.02% NaN₃ in a small, air-tight plastic box 5 overnight in a cold room. The next day streptavidin is removed and replaced with at least 10 ml blocking solution (29 mg/ml of BSA; 3 µg/ml of streptavidin; 0.1 M NaHCO₃ pH 8.6-0.02% NaN₃) and incubated at least 1 hour at room 10 temperature. The blocking solution is removed and plates are washed rapidly three times with Tris buffered saline containing 0.5% Tween 20 (TBS-0.5% Tween 20).

Selection of phage expressing peptides bound by the ligand binding proteins is performed with 5 µl (2.7 µg ligand binding protein) of blocked biotinylated ligand 15 binding proteins reacted with a 50 µl portion of each library. Each mixture is incubated overnight at 4°C, diluted with 1 ml TBS-0.5% Tween 20, and transferred to a streptavidin-coated petri plate prepared as described above. After rocking 10 minutes at room temperature, 20 unbound phage are removed and plates washed ten times with TBS-0.5% Tween 20 over a period of 30-90 minutes. Bound phage are eluted from plates with 800 µl sterile elution buffer (1 mg/ml BSA, 0.1 M HCl, pH adjusted to 2.2 with glycerol) for 15 minutes and eluates neutralized with 48 µl 25 2 M Tris (pH unadjusted). A 20 µl portion of each eluate is titered on MK30-3 concentrated cells with dilutions of input phage.

A second round of panning is performed by treating 750 µl of first eluate from each library with 5 mM DTT for 10 30 minutes to break disulfide bonds linking biotin groups to residual biotinylated binding proteins. The treated eluate is concentrated on a Centricon 30 ultrafilter (Amicon), washed three times with TBS-0.5% Tween 20, and concentrated to a final volume of about 50 µl. Final retentate is 35 transferred to a tube containing 5.0 µl (2.7 µg ligand

binding protein) blocked biotinylated ligand binding proteins and incubated overnight. The solution is diluted with 1 ml TBS-0.5% Tween 20, panned, and eluted as described above on fresh streptavidin-coated petri plates.

5 The entire second eluate (800 μ l) is neutralized with 48 μ l 2 M Tris, and 20 μ l is titered simultaneously with the first eluate and dilutions of the input phage.

Individual phage populations are purified through 2 to 3 rounds of plaque purification. Briefly, the second 10 eluate titer plates are lifted with nitrocellulose filters (Schleicher & Schuell, Inc., Keene, NH) and processed by washing for 15 minutes in TBS (10 mM Tris-HCl, pH 7.2, 150 mM NaCl), followed by an incubation with shaking for an additional 1 hour at 37°C with TBS containing 5% nonfat dry 15 milk (TBS-5% NDM) at 0.5 ml/cm². The wash is discarded and fresh TBS-5% NDM is added (0.1 ml/cm²) containing the ligand binding protein between 1 nM to 100 mM, preferably between 1 to 100 μ M. All incubations are carried out in heat- 20 sealable pouches (Sears). Incubation with the ligand binding protein proceeds for 12-16 hours at 4°C with shaking. The filters are removed from the bags and washed 3 times for 30 minutes at room temperature with 150 mls of TBS containing 0.1% NDM and 0.2% NP-40 (Sigma, St. Louis, MO). The filters are then incubated for 2 hours at room 25 temperature in antiserum against the ligand binding protein at an appropriate dilution in TBS-0.5% NDM, washed in 3 changes of TBS containing 0.1% NDM and 0.2% NP-40 as described above and incubated in TBS containing 0.1% NDM and 0.2% NP-40 with 1 x 10⁶ cpm of ¹²⁵I-labeled Protein A 30 (specific activity = 2.1 x 10⁷ cpm/ μ g). After a washing with TBS containing 0.1% NDM and 0.2% NP-40 as described above, the filters are wrapped in Saran Wrap and exposed to Kodak X-Omat x-ray film (Kodak, Rochester, NY) for 1-12 hours at -70°C using Dupont Cronex Lightning Plus 35 Intensifying Screens (Dupont, Willmington, DE).

Positive plaques identified are cored with the large end of a pasteur pipet and placed into 1 ml of SM (5.8 g NaCl, 2 g MgSO₄·7H₂O, 50 ml 1 M Tris-HCl, pH 7.5, 5 mls 2% gelatin, to 1000 mls with dH₂O) plus 1-3 drops of CHCl₃ and 5 incubated at 37°C 2-3 hours or overnight at 4°C. The phage are diluted 1:500 in SM and 2 µl are added to 300 µl of XL1 cells plus 3 mls of soft agar per 100 mm² plate. The XL1 cells are prepared for plating by growing a colony overnight in 10 ml LB (10 g bacto-tryptone, 5 g bacto-yeast extract, 10 g NaCl, 1000 ml dH₂O) containing 100 µl of 20% maltose and 100 µl of 1 M MgSO₄. The bacteria are pelleted by centrifugation at 2000 xg for 10 minutes and the pellet is resuspended gently in 10 mls of 10 mM MgSO₄. The suspension is diluted 4-fold by adding 30 mls of 10 mM MgSO₄ 10 to give an OD₆₀₀ of approximately 0.5. The second and third round screens are identical to that described above except that the plaques are cored with the small end of a pasteur pipet and placed into 0.5 mls SM plus a drop of CHCl₃ and 1- 15 µl of the phage followin, incubation are used for plating 5 µl of the phage without dilution. At the end of the third round of 20 purification, an individual plaque is picked and the templates prepared for sequencing.

Template Preparation and Sequencing

Templates are prepared for sequencing by inoculating 25 a 1 ml culture of 2XYT containing a 1:100 dilution of an overnight culture of XL1 with an individual plaque. The plaques are picked using a sterile toothpick. The culture is incubated at 37°C for 5-6 hours with shaking and then transferred to a 1.5 ml microfuge tube. 200 µl of PEG 30 solution is added, followed by vortexing and placed on ice for 10 minutes. The phage precipitate is recovered by centrifugation in a microfuge at 12,000 x g for 5 minutes. The supernatant is discarded and the pellet is resuspended in 230 µl of TE (10 mM Tris-HCl, pH 7.5, 1 mM EDTA) by 35 gently pipeting with a yellow pipet tip. Phenol (200 µl)

is added, followed by a brief vortex and microfuged to separate the phases. The aqueous phase is transferred to a separate tube and extracted with 200 μ l of phenol/chloroform (1:1) as described above for the phenol extraction. A 0.1 volume of 3 M NaOAc is added, followed by addition of 2.5 volumes of ethanol and precipitated at -20°C for 20 minutes. The precipitated templates are recovered by centrifugation in a microfuge at 12,000 x g for 8 minutes. The pellet is washed in 70% ethanol, dried and resuspended in 25 μ l TE. Sequencing was performed using a Sequenase™ sequencing kit following the protocol supplied by the manufacturer (U.S. Biochemical, Cleveland, OH).

EXAMPLE II

15 Isolation and Characterization of Peptide Ligands Generated
From Oligonucleotides Having Random Codons at Two
Predetermined Positions

This example shows the generation of a surface expression library from a population of oligonucleotides having randomized codons. The oligonucleotides are ten codons in length and are cloned into a single vector species for the generation of a M13 gene VIII-based surface expression library. The example also shows the selection of peptides for a ligand binding protein and characterization of their encoded nucleic acid sequences.

Oligonucleotide Synthesis

Oligonucleotides were synthesized as described in Example I. The synthesizer was programmed to synthesize the sequences shown in Table IX. These sequences correspond to the first random codon position synthesized and 3' flanking sequences of the oligonucleotide which hybridizes to the leader sequence in the vector. The

complementary sequences are used for insertional mutagenesis of the synthesized population of oligonucleotides.

Table IX

	<u>Column</u>	<u>Sequence (5' to 3')</u>
5	column 1	AA(A/C)GGTTGGTCGGTACCGG
	column 2	AG(A/G)GGTTGGTCGGTACCGG
	column 3	AT(A/G)GGTTGGTCGGTACCGG
	column 4	AC(A/G)GGTTGGTCGGTACCGG
10	column 5	CA(G/T)GGTTGGTCGGTACCGG
	column 6	CT(G/C)GGTTGGTCGGTACCGG
	column 7	AG(T/C)GGTTGGTCGGTACCGG
	column 8	AT(T/C)GGTTGGTCGGTACCGG
	column 9	CC(A/C)GGTTGGTCGGTACCGG
15	column 10	T(A/T)TGGTTGGTCGGTACCGG

The next eight random codon positions were synthesized as described for Table V in Example I. Following the ninth position synthesis, the reaction products were once more combined, mixed and redistributed into 10 new reaction 20 columns. Synthesis of the last random codon position and 5' flanking sequences are shown in Table X.

Table X

	<u>Column</u>	<u>Sequence (5' to 3')</u>
25	column 1	AGGATCCGCCGAGCTCAA(A/C)A
	column 2	AGGATCCGCCGAGCTCAG(A/G)A
	column 3	AGGATCCGCCGAGCTCAT(A/G)A
	column 4	AGGATCCGCCGAGCTCAC(A/G)A
	column 5	AGGATCCGCCGAGCTCCA(G/T)A
	column 6	AGGATCCGCCGAGCTCCT(G/C)A
30	column 7	AGGATCCGCCGAGCTCAG(T/C)A
	column 8	AGGATCCGCCGAGCTCAT(T/C)A
	column 9	AGGATCCGCCGAGCTCCC(A/C)A
	column 10	AGGATCCGCCGAGCTCT(A/T)TA

The reaction products were mixed once more and the oligonucleotides cleaved and purified as recommended by the manufacturer. The purified population of oligonucleotides were used to generate a surface expression library as 5 described below.

Vector Construction

The vector used for generating surface expression libraries from a single oligonucleotide population (i.e., without joining together of right and left half oligonucleotides) is described below. The vector is a M13-based expression vector which directs the synthesis of gene VIII-peptide fusion proteins (Figure 4). This vector exhibits all the functions that the combined right and left half vectors of Example I exhibit.

An M13-based vector was constructed for the cloning and surface expression of populations of random oligonucleotides (Figure 4, M13IX30), M13mp19 (Pharmacia) was the starting vector. This vector was modified to contain, in addition to the encoded wild type M13 gene 15 VIII: (1) a pseudo-wild type gene, gene VIII sequence with an amber stop codon placed between it and the restriction sites for cloning oligonucleotides; (2) Stu I, Spe I and Xho I restriction sites in frame with the pseudo-wild type gVIII for cloning oligonucleotides; (3) sequences necessary 20 for expression, such as a promoter, signal sequence and translation initiation signals; (4) various other mutations 25 to remove redundant restriction sites and the amino terminal portion of Lac Z.

Construction of M13IX30 was performed in four steps. 30 In the first step, a precursor vector containing the pseudo gene VIII and various other mutations was constructed, M13IX01F. The second step involved the construction of a small cloning site in a separate M13mp18 vector to yield

M13IX03. In the third step, expression sequences and cloning sites were constructed in M13IX03 to generate the intermediate vector M13IX04B. The fourth step involved the incorporation of the newly constructed sequences from the 5 intermediate vector into M13IX01F to yield M13IX30. Incorporation of these sequences linked them with the pseudo gene VIII.

Construction of the precursor vector M13IX01F was similar to that of M13IX42 described in Example I except for the following features: (1) M13mp19 was used as the starting vector; (2) the Fok I site 5' to the unique Eco RI site was not incorporated and the overhang at the naturally occurring Fok I site at position 3547 was not changed to 5'-CTTC-3'; (3) the spacer sequence was not incorporated between the Eco RI and Sac I sites; and (4) the amber codon at position 4492 was not incorporated.

In the second step, M13mp18 was mutated to remove the 5' end of Lac Z up to the Lac i binding site and including the Lac Z ribosome binding site and start codon. Additionally, the polylinker was removed and a Mlu I site was introduced in the coding region of Lac Z. A single oligonucleotide was used for these mutagenesis and had the sequence "5'-AAACGACGGCCAGTGCCAAGTGACGCGTGTGAAATTGTTATCC-3'" (SEQ ID NO: 41). Restriction enzyme sites for Hind III and Eco RI were introduced downstream of the MluI site 25 using the oligonucleotide "5'-GGCGAAAGGAAATTCTGCAAGGCGATTAAGCTTGGTAACGCC-3'" (SEQ ID NO: 42). These modifications of M13mp18 yielded the vector M13IX03.

30 The expression sequences and cloning sites were introduced into M13IX03 by chemically synthesizing a series of oligonucleotides which encode both strands of the desired sequence. The oligonucleotides are presented in Table XI (SEQ ID NOS: 43 through 50).

TABLE XI
M13IX30 Oligonucleotide Series

<u>Top Strand Oligonucleotides</u>		<u>Sequence (5' to 3')</u>
5	084	GGCGTTACCCAAGCTTGACATGGAGAAAATAAAG
	027	TGAAACAAAGCACTATTGCACTGGCACTCTTACCGT TACCGT
	028	TACTGTTACCCCTGTGACAAAAGCCGCCAGGTCC AGCTGC
10	029	TCGAGTCAGGCCTATTGTGCCAGGGATTGTACTAG TGGATCCG
<u>Bottom Oligonucleotides</u>		<u>Sequence (5' to 3')</u>
15	085	TGGCGAAAGGGAATTGGATCCACTAGTACAATCCCTG
	031	GGCACAAATAGGCCTGACTCGAGCAGCTGGACCAGGGCG GCTT
	032	TTGTCACAGGGTAAACAGTAACGGTAACGGTAAGTGT GCCA
	033	GTGCAATAGTGCTTGTTCACTTATTTCTCCATGT ACAA

The above oligonucleotides except for the terminal oligonucleotides 084 (SEQ ID NO: 43) and 085 (SEQ ID NO: 47) of Table XI were mixed, phosphorylated, annealed and ligated to form a double stranded insert as described in Example I. However, instead of cloning directly into the intermediate vector the insert was first amplified by PCR using the terminal oligonucleotides 084 (SEQ ID NO: 43) and 085 (SEQ ID NO: 47) as primers. The terminal oligonucleotide 084 (SEQ ID NO: 43) contains a Hind III site 10 nucleotides internal to its 5' end. Oligonucleotide 085 (SEQ ID NO: 47) has an Eco RI site at its 5' end. Following amplification, the products were restricted with Hind III and Eco RI and ligated as described in Example I into the polylinker of M13mp18

digested with the same two enzymes. The resultant double stranded insert contained a ribosome binding site, a translation initiation codon followed by a leader sequence and three restriction enzyme sites for cloning random 5 oligonucleotides (Xho I, Stu I, Spe I). The vector was named M13IX04.

During cloning of the double-stranded insert, it was found that one of the GCC codons in oligonucleotides 028 and its complement in 031 was deleted. Since this deletion 10 did not affect function, the final construct is missing one of the two GCC codons. Additionally, oligonucleotide 032 contained a GTG codon where a GAG codon was needed. Mutagenesis was performed using the oligonucleotide 5'-TAACGGTAAGAGTGCCAGTGC-3' (SEQ ID NO: 51) to convert the 15 codon to the desired sequence. The resultant intermediate vector was named M13IX04B.

The fourth step in constructing M13IX30 involved inserting the expression and cloning sequences from M13IX04B upstream of the pseudo-wild type gVIII in 20 M13IX01F. This was accomplished by digesting M13IX04B with Dra III and Ban HI and gel isolating the 700 base pair insert containing the sequences of interest. M13IX01F was likewise digested with Dra III and Bam HI. The insert was combined with the double digested vector at a molar ratio 25 of 3:1 and ligated as described in Example I. It should be noted that all modifications in the vectors described herein were confirmed by sequence analysis. The sequence of the final construct, M13IX30, is shown in Figure 7 (SEQ ID NO: 3). Figure 4 also shows M13IX30 where each of the 30 elements necessary for surface expression of randomized oligonucleotides is marked.

Library Construction, Screening and Characterization of
Encoded Oligonucleotides

Construction of an M13IX30 surface expression library is accomplished identically to that described in Example I for sublibrary construction except the oligonucleotides described above are inserted into M13IX30 by mutagenesis instead of by ligation. The library is constructed and propagated on MK30-3 (BMB) and phage stocks are prepared for infection of XLI cells and screening. The surface expression library is screened and encoding oligonucleotides characterized as described in Example I.

EXAMPLE III

Isolation and Characterization of Peptide Ligands
Generated from Right and Left Half
Degenerate Oligonucleotides

15

This example shows the construction and expression of a surface expression library of degenerate oligonucleotides. The encoded peptides of this example derive from the mixing and joining together of two 20 separate oligonucleotide populations. Also demonstrated is the isolation and characterization of peptide ligands and their corresponding nucleotide sequence for specific binding proteins.

Synthesis of Oligonucleotide Populations

25 A population of left half degenerate oligonucleotides and a population of right half degenerate oligonucleotides was synthesized using standard automated procedures as described in Example I.

30 The degenerate codon sequences for each population of oligonucleotides were generated by sequentially

synthesizing the triplet NNG/T where N is an equal mixture of all four nucleotides. The antisense sequence for each population of oligonucleotides was synthesized and each population contained 5' and 3' flanking sequences complementary to the vector sequence. The complementary termini was used to incorporate each population of oligonucleotides into their respective vectors by standard mutagenesis procedures. Such procedures have been described previously in Example I and in the Detailed Description. Synthesis of the antisense sequence of each population was necessary since the single-stranded form of the vectors are obtained only as the sense strand.

The left half oligonucleotide population was synthesized having the following sequence: 5'-AGCTCCCGGATGCCTCAGAAGATG(A/CNN)₉GGCTTTGCCACAGGGG-3' (SEQ ID NO: 52). The right half oligonucleotide population was synthesized having the following sequence: 5'-CAGCCTCGGATCCGCC(A/CNN)₁₀ATG(A/C)GAAT-3' (SEQ ID NO. 53). These two oligonucleotide populations when incorporated into their respective vectors and joined together encode a 20 codon oligonucleotide having 19 degenerate positions and an internal predetermined codon sequence.

Vector Construction

Modified forms of the previously described vectors were used for the construction of right and left half sublibraries. The construction of left half sublibraries was performed in an M13-based vector termed M13ED03. This vector is a modified form of the previously described M13IX30 vector and contains all the essential features of both M13IX30 and M13IX22. M13ED03 contains, in addition to a wild type and a pseudo-wild type gene VIII, sequences necessary for expression and two Fok I sites for joining with a right half oligonucleotide

sublibrary. Therefore, this vector combines the advantages of both previous vectors in that it can be used for the generation and expression of surface expression libraries from a single oligonucleotide 5 population or it can be joined with a sublibrary to bring together right and left half oligonucleotide populations into a surface expression library.

M13ED03 was constructed in two steps from M13IX30. The first step involved the modification of M13IX30 to 10 remove a redundant sequence and to incorporate a sequence encoding the eight amino-terminal residues of human β -endorphin. The leader sequence was also mutated to increase secretion of the product.

During construction of M13IX04 (an intermediate 15 vector to M13IX30 which is described in Example II), a six nucleotide sequence was duplicated in oligonucleotide 027 (SEQ ID NO: 44) and its complement 032 (SEQ ID NO: 49). This sequence, 5'-TTACCG-3', was deleted by mutagenesis in the construction of M13ED01. The 20 oligonucleotide used for the mutagenesis was 5'-GGTAAACAGTAACGGTAAGAGTGCCAG-3' (SEQ ID NO: 54). The mutation in the leader sequence was generated using the oligonucleotide 5'-GGGCTTTGCCACAGGGGT-3' (SEQ ID NO: 55). This mutagenesis resulted in the A residue at 25 position 6353 of M13IX30 being changed to a G residue. The resultant vector was designated M13IX32.

To generate M13ED01, the nucleotide sequence 30 encoding β -endorphin (8 amino acid residues of β -endorphin plus 3 extra amino acid residues) was incorporated after the leader sequence by mutagenesis. The oligonucleotide used had the following sequence: 5'-AGGGTCATGCCCTTCAGCTCCGGATCCCTCAGAAGTCATAAACCCCCCATAGGC 35 TTTTGCAC-3' (SEQ ID NO: 56). This mutagenesis also removed some of the downstream sequences through the Spe

I site.

The second step in the construction of M13ED03 involved vector changes which put the β -endorphin sequence in frame with the downstream pseudo-gene VIII sequence and incorporated a Fok I site for joining with a sublibrary of right half oligonucleotides. This vector was designed to incorporate oligonucleotide populations by mutagenesis using sequences complementary to those flanking or overlapping with the encoded β -endorphin sequence. The absence of β -endorphin expression after mutagenesis can therefore be used to measure the mutagenesis frequency. In addition to the above vector changes, M13ED03 was also modified to contain an amber codon at position 3262 for biological selection during joining of right and left half sublibraries.

The mutations were incorporated using standard mutagenesis procedures as described in Example I. The frame shift changes and Fok I site were generated using the oligonucleotide 5'-
20 TCGCCTTCAGCTCCGGATGCCTCAGAAGCATGAACCCCCATAGGC-3' (SEQ ID NO: 57). The amber codon was generated using the oligonucleotide 5'-CAATTATCCTAAATCTTACCAAC-3' (SEQ ID NO: 58). The full sequence of the resultant vector, M13ED03, is provided in Figure 8 (SEQ ID NO: 4).

25 The construction of right half oligonucleotide sublibraries was performed in a modified form of the M13IX42 vector. The new vector, M13IX421, is identical to M13IX42 except that the amber codon between the Eco RI-SacI cloning site and the pseudo-gene VIII sequence 30 was removed. This change ensures that all expression off of the Lac Z promoter produces a peptide-gene VIII fusion protein. Removal of the amber codon was performed by mutagenesis using the following oligonucleotide: 5'- GCCTTCAGCCTCGGATCCGCC-3' (SEQ ID NO: 59). The full

sequence of M13IX421 is shown in Figure 9 (SEQ ID NO: 5).

Library Construction, Screening and Characterization of
Encoded Oligonucleotides

A sublibrary was constructed for each of the
5 previously described degenerate populations of
oligonucleotides. The left half population of
oligonucleotides was incorporated into M13ED03 to
generate the sublibrary M13ED03.L and the right half
population of oligonucleotides was incorporated into
10 M13IX421 to generate the sublibrary M13IX421.R. Each of
the oligonucleotide populations were incorporated into
their respective vectors using site-directed mutagenesis
as described in Example I. Briefly, the nucleotide
sequences flanking the degenerate codon sequences were
15 complementary to the vector at the site of incorporation.
The populations of nucleotides were hybridized to single-
stranded M13ED03 or M13IX421 vectors and extended with T4
DNA polymerase to generate a double-stranded circular
vector. Mutant templates were obtained by uridine
20 selection *in vivo*, as described by Kunkel et al., supra.
Each of the vector populations were electroporated into
host cells and propagated as described in Example I.

The random joining of right and left half
sublibraries into a single surface expression library was
25 accomplished as described in Example I except that prior
to digesting each vector population with Fok I they were
first digested with an enzyme that cuts in the unwanted
portion of each vector. Briefly, M13ED03.L was digested
with Bgl II (cuts at 7094) and M13IX421.R was digested
30 with Hind III (cuts at 3919). Each of the digested
populations were further treated with alkaline
phosphatase to ensure that the ends would not religate
and then digested with an excess of Fok I. Ligations,
electroporation and propagation of the resultant library

was performed as described in Example I.

The surface expression library was screened for ligand binding proteins using a modified panning procedure. Briefly, 1 ml of the library, about 10^{12} phage particles, was added to 1-5 μ g of the ligand binding protein. The ligand binding protein was either an antibody or receptor globulin (Rg) molecule, Aruffo et al., Cell 61:1303-1313 (1990), which is incorporated herein by reference. Phage were incubated shaking with affinity ligand at room temperature for 1 to 3 hours followed by the addition of 200 μ l of latex beads (Biosite, San Diego, CA) which were coated with goat-antimouse IgG. This mixture was incubated shaking for an additional 1-2 hours at room temperature. Beads were pelleted for 2 minutes by centrifugation in a microfuge and washed with TBS which can contain 0.1% Tween 20. Three additional washes were performed where the last wash did not contain any Tween 20. The bound phage were then eluted with 200 μ l 0.1 M Glycine-HCl, pH 2.2 for 15 minutes and the beads were spun down by centrifugation. The supernatant-containing phage (eluate) was removed and phage exhibiting binding to the ligand binding protein were further enriched by one-to-two more cycles of panning. Typical yields after the first eluate were about 1×10^6 - 5×10^6 pfu. The second and third eluate generally yielded about 5×10^6 - 2×10^7 pfu and 5×10^7 - 1×10^{10} pfu, respectively.

The second or third eluate was plated at a suitable density for plaque identification screening and sequencing of positive clones (i.e., plated at confluence for rare clones and 200-500 plaques/plate if pure plaques were needed). Briefly, plaques grown for about 6 hours at 37°C and were overlaid with nitrocellulose filters that had been soaked in 2 mM IPTG and then briefly dried. The filters remained on the plaques overnight at room

temperature, removed and placed in blocking solution for 1-2 hours. Following blocking, the filters were incubated in 1 μ g/ml ligand binding protein in blocking solution for 1-2 hours at room temperature. Goat 5 antimouse Ig-coupled alkaline phosphatase (Fisher) was added at a 1:1000 dilution and the filters were rapidly washed with 10 mls of TBS or block solution over a glass vacuum filter. Positive plaques were identified after alkaline phosphatase development for detection.

10 Screening of the degenerate oligonucleotide library with several different ligand binding proteins resulted in the identification of peptide sequences which bound to each of the ligands. For example, screening with an antibody to β -endorphin resulted in the detection of 15 about 30-40 different clones which essentially all had the core amino acid sequence known to interact with the antibody. The sequences flanking the core sequences were different showing that they were independently derived and not duplicates of the same clone. Screening with an 20 antibody known as 57 gave similar results (i.e., a core consensus sequence was identified but the flanking sequences among the clones were different).

EXAMPLE IV

Generation of a Left Half Random Oligonucleotide Library

25 This example shows the synthesis and construction of a left half random oligonucleotide library.

A population of random oligonucleotides nine codons in length was synthesized as described in Example I except that different sequences at their 5' and 3' ends 30 were synthesized so that they could be easily inserted into the vector by mutagenesis. Also, the mixing and dividing steps for generating random distributions of

reaction products was performed by the alternative method of dispensing equal volumes of bead suspensions. The liquid chosen that was dense enough for the beads to remain dispersed was 100% acetonitrile.

5 Briefly, each column was prepared for the first coupling reaction by suspending 22 mg (1 μ mole) of 48 μ mol/g capacity beads (Genta, San Diego, CA) in 0.5 mls of 100% acetonitrile. These beads are smaller than those described in Example I and are derivatized with a guanine 10 nucleotide. They also do not have a controlled pore size. The bead suspension was then transferred to an empty reaction column. Suspensions were kept relatively dispersed by gently pipetting the suspension during transfer. Columns were plugged and monomer coupling 15 reactions were performed as shown in Table XII.

Table XII

	<u>Column</u>	<u>Sequence (5' to 3')</u>
20	column 1L	AA(A/C)GGCTTTGCCACAGG
	column 2L	AG(A/G)GGCTTTGCCACAGG
	column 3L	AT(A/G)GGCTTTGCCACAGG
	column 4L	AC(A/G)GGCTTTGCCACAGG
	column 5L	CA(G/T)GGCTTTGCCACAGG
	column 6L	CT(G/C)GGCTTTGCCACAGG
25	column 7L	AG(T/C)GGCTTTGCCACAGG
	column 8L	AT(T/C)GGCTTTGCCACAGG
	column 9L	CC(A/C)GGCTTTGCCACAGG
	column 10L	T(A/T)TGGCTTTGCCACAGG

After coupling of the last monomer, the columns were 30 unplugged as described previously and their contents were poured into a 1.5 ml microfuge tube. The columns were rinsed with 100% acetonitrile to recover any remaining beads. The volume used for rinsing was determined so

that the final volume of total bead suspension was about 100 μ l for each new reaction column that the beads would be aliquoted into. The mixture was vortexed gently to produce a uniformly dispersed suspension and then 5 divided, with constant pipetting of the mixture, into equal volumes. Each mixture of beads was then transferred to an empty reaction column. The empty tubes were washed with a small volume of 100% acetonitrile and also transferred to their respective columns. Random 10 codon positions 2 through 9 were then synthesized as described in Example I where the mixing and dividing steps were performed using a suspension in 100% acetonitrile. The coupling reactions for codon positions 2 through 9 are shown in Table XIII.

15

Table XIII

<u>Column</u>	<u>Sequence (5' to 3')</u>
column 1L	AA(A/C)A
column 2L	AG(A/G)A
20 column 3L	AT(A/G)A
column 4L	AC(A/G)A
column 5L	CA(G/T)A
column 6L	CT(G/C)A
column 7L	AG(T/C)A
25 column 8L	AT(T/C)A
column 9L	CC(A/C)A
column 10L	T(A/T)TA

After coupling of the last monomer for the ninth codon position, the reaction products were mixed and a 30 portion was transferred to an empty reaction column. Columns were plugged and the following monomer coupling reactions were performed: 5'-CGGATGCCTCAGAAGCCCCXXA-3' (SEQ ID NO: 60). The resulting population of random oligonucleotides was purified and incorporated by

mutagenesis into the left half vector M13ED04.

M13ED04 is a modified version of the M13ED03 vector described in Example III and therefore contains all the features of that vector. The difference between M13ED03 and M13ED04 is that M13ED04 does not contain the five amino acid sequence (Tyr Gly Gly Phe Met) recognized by anti- β -endorphin antibody. This sequence was deleted by mutagenesis using the oligonucleotide 5'-
10 CGGATGCCTCAGAAGGGCTTTGCCACAGG (SEQ ID NO: 61). The entire nucleotide sequence of this vector is shown in Figure 10 (SEQ ID NO: 6).

Although the invention has been described with reference to the presently preferred embodiment, it should be understood that various modifications can be made without departing from the spirit of the invention. Accordingly, the invention is limited only by the claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Huse, William D.
- (ii) TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF RANDOMIZED PEPTIDES
- (iii) NUMBER OF SEQUENCES: 61
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
 - (B) STREET: 444 South Flower Street, Suite 2000
 - (C) CITY: Los Angeles
 - (D) STATE: California
 - (E) COUNTRY: United States
 - (F) ZIP: 90071
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Campbell, Cathryn A
 - (B) REGISTRATION NUMBER: 31,815
 - (C) REFERENCE/DOCKET NUMBER: P31 9072
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (619) 535-9001
 - (B) TELEFAX: (619) 535-8949

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 7294 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTTCAG CTCGGCCCC AAATGAAAAT	60
ATAGCTAAC AGGTTATTGA CCATTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT	120
CGTTCCGAGA ATTGGGAATC AACTGTTACA TCGAATGAAA CTTCCAGACA CCGTACTTTA	180
GTTGCATATT TAAAACATCT TGAGCTACAG CACCAAGATTG AGCAATTAAG CTCTAAGCCA	240
TCTGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG	300
TTGGAGTTG CTTCCGGTCT GGTTCGCTT GAAGCTCGAA TTAAAACGCG ATATTTGAAG	360
TCTTTCGGGC TTCCTCTTAA TCTTTTGAT GCAATCCGCT TTGCTTCTGA CTATAATAGT	420

CAGGGTAAAG ACCTGATTIT TGATITATGG TCATTCTCGT TTTCTGAAC	450
TTTGAGGGGG ATTCAATGAA TATTATGAC GATTCCGCAG TATTGGACGC	540
AAACATTTA CTATTACCCC CTCTGGCAAA ACTTCTTTG CAAAAGCCTC	600
GGTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTAC	660
AATTCCCTTT CCCGTTATGT ATCTGCATTA GTTGAATGTG GTATTCTAA	720
ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA	780
TCTTCCCAAC GTCCTGACTG GTATAATGAG CCAGTTCTTA AAATCGCATA	840
CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT	900
CTCGTCAGGG CAAGCCTAT TCACTGAATG AGCAGCTTG TTACGTTGAT	960
AATATCCGGT TCTTGTCAAG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT	1020
TGTACACCGT TCATCTGTCC TCTTCAAAG TTGGTCAGTT CGGTTCCCTT	1080
GTCTGCCCT CGTTCCGGCT AAGTAACATG GAGCAGGTG CGGATTTCGA	1140
CAGGCATGA TACAAATCTC CGTTGTACTT TGTTTCGGC TTGGTATAAT	1200
CAAAGATGAG TGTTTTAGTG TATTCTTCG CCTCTTGTGTT TAGGTTGG	1260
GTGGCATTAC GTATTTTACCGT CGTTAATGG AAACCTTCCTC ATGAAAAAGT	1320
CAAAGCCTCT GTAGCCGTG CTACCCCTCGT TCCGATGCTG TCTTCGCTG	1380
CGATCCCGCA AAAGCGGGCT TTAACCTCCCT GCAAGCCTCA GCGACCGAAT	1440
TCCGTGGCG ATGGTTGTG TCATTGTCGG CGCAACTATC GGTATCAAGC	1500
ATTCACCTCG AAAGGAAGCT GATAAACCGA TACAATTAAA GGCTCCTTTT	1560
TTTTGGAGA TTTTCAACGT GAAAAAATTA TTATTGCAA TTGCTTTAGT	1620
TATTCTCACT CGCGTAAAC TGTGAAAGT TGTGTTAGCAA AACCCCATAC	1680
TTTACTAACG TCTGGAAAGA CGACAAAAGT TTAGATCGTT ACGCTAACTA	1740
CTGTGGAATG CTACAGGCGT TGTAGTTGT ACTGGTGAGG AAACCTCAGTG	1800
TGGGTTCTA TTGGGCTTGC TATCCCTGAA AATGAGGGTG GTGGCTCTGA	1860
TCTGAGGGTG GCGGTTCTGA GGGTGGCGGT ACTAAACCCTC CTGAGTACGG	1920
ATTCCGGGCT ATACTTATAT CAACCCCTCTC GACGGCACTT ATCCGCCTGG	1980
AACCCCGCTA ATCCTAATCC TTCTCTTGAG GAGTCTCAGC CTCTTAATAC	2040
CAGAATAATA GGTTCCGAAA TAGGCAGGGG GCATTAACGT TTTATACGGG	2100
CAAGGCAC TG ACCCCGTTAA AACTTATTAC CAGTACACTC CTGTATCATC	2160
TATGACGCTT ACTGGAACGG TAAATTCAAGA GACTGGCTT TCCATTCTGG	2220
GATCCATTGCG TTTGTGAATA TCAAGGCCAA TCGTCTGACC TGCTCAACC	2280
GCTGGCGGGC GCTCTGGTGG TGTTCTGGT GGCGGCTCTG AGGGTGGTGG	2340
GGCGGTTCTG AGGGTGGCGG CTCTGAGGGA GGCGGTTCCG GTGGTGGCTC	2400
GATTTTGATT ATGAAAAGAT GCCAAACGCT AATAAGGGGG CTATGACCGA	2460

GAAAACGCGC TACAGTCTGA CGCTAAAGGC AAACTTGATT CTGTCGCTAC TGATTACGGT	3520
GCTGCTATCG ATGGTTCAT TGGTGACGTT TCCGGCCTTG CTAATGGTAA TGGTGCTACT	3580
GGTGATTTG CTGGCTCTAA TTCCCAAATG GCTCAAGTCG GTGACGGTGA TAATTACCT	3640
TTAATGAATA ATTTCCGTCA ATATTTACCT TCCCTCCCTC AATCGGTTGA ATGTCGCCCT	3700
TTTGTCTTTA GCGCTGGTAA ACCATATGAA TTTTCTATTG ATTGTGACAA AATAAACTTA	3760
TTCCGTGGTG TCTTTGCGTT TCTTTATAT GTTGCACCT TTATGTATGT ATTTTCTACG	3820
TTTGCTAAC A TACTGCGTAA TAAGGAGTCT TAATCATGCC AGTTCTTTG GGTATTCCGT	3880
TATTATTGCG TTTCCCTCGGT TTCCCTCTGG TAACTTGTT CGGCTATCTG CTTACTTTTC	3940
TTAAAAAGGG CTTCCGTAAG ATAGCTATTG CTATTCATT GTTCTTGCT CTTATTATTG	3000
GGCTTAACTC AATTCTTGTC GGTTATCTCT CTGATATTAG CGCTCAATTAA CCCTCTGACT	3060
TTGTTCAAGGG TGTTCAAGTTA ATTCTCCCGT CTAATGCGCT TCCCTGTTT TATGTTATTG	3120
TCTCTGTAAA GGCTGCTATT TTCATTTTG ACGTTAAACA AAAAATCGTT TCTTATTG	3180
ATTGGGATAA ATAATATGGC TGTTTATTTT GTAACTGGCA AATTAGGCTC TCGAAAGACG	3240
CTCGTTAGCG TTGGTAAGAT TCAGGATAAA ATTGTAGCTG GGTGAAAAT AGCAACTAAT	3300
CTTGATTTAA GGCTCAAAA CCTCCCGCAA GTCGGAGGT TCGCTAAAAC CCCTCGCGTT	3360
CTTAGAATAC CGGATAAGCC TTCTATATCT GATTTGCTTG CTATTGGCG CGGTAAATGAT	3420
TCCTACGATG AAAATAAAA CGGCTTGCTT GTTCTCGATG AGTGGGTAC TTGGTTAAAT	3480
ACCCGTTCTT CGAATGATAA GGAAAGACAG CGGATTATTG ATTGGTTCT ACATGCTCGT	3540
AAATTAGGAT GGGATATTAT CTTCCTTGTT CAGGACTTAT CTATTGTTGA TAAACAGGGC	3600
GGTTCTGCAT TAGCTGAACA TGTTGTTAT TGTCGTGGC TGGACAGAAAT TACTTTACCT	3660
TTTGTGGTA CTTTATATTG TCTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT	3720
GTTGGCGTTG TTAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGGG TTGGCTTTAT	3780
ACTGGTAAGA ATTTGTATAA CGCATATGAT ACTAACAGG CTTTTCTAG TAATTATGAT	3840
TCCGGTGTAT ATTCTTATTT AACGCCCTAT TTATCACACG GTCGGTATTT CAAACCATTAA	3900
AATTAGGTC AGAAGATGAA GCTTACTAAA ATATATTGA AAAAGTTTTC ACGCGTTCTT	3960
TGTCTTGGCA TTGGATTGTC ATCAGCATT ACATATAGTT ATATAACCCA ACCTAAGCCG	4020
GAGGTTAAA AGGTAGTCTC TCAGACCTAT GATTTGATA AATTCACTAT TGACTCTCT	4080
CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAAAT	4140
AGCGACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTTATG TACTGTTCC	4200
ATTAAAAAGG TAATTCAAAT GAAATTGTTA AATGTAATTAA ATTTGTTTT CTTGATGTTT	4260
GTTTCATCAT CTTCTTTGC TCAGGTAATT GAAATGAATA ATTCGCCTCT GCGCGATTTC	4320
GTAACCTGGT ATTCAAAGCA ATCAGGCCAA TCCGTTATTG TTTCTCCGA TGTAAAAGGT	4380
ACTGTTACTG TATATTCACTC TGACGTTAAA CCTGAAAATC TACGCAATTCTTATTCT	4440
GTTTTACGTG CTAATAATTG TGATATGGTT GGTCGAATTG CTTCCATTAT TTAGAAGTAT	4500

AATCCAAACA ATCAGGATTA TATTGATGAA TTGCCATCAT CTGATAATCA GGAATATGAT	4560
GATAATTCCG CTCCCTCTGG TGGTTCTTT GTTCCGGAAA ATGATAATGT TACTCAAAC	4620
TTTAAAATTA ATAACGTTCG GGCAAAGGAT TTAATACGAG TTGTCGAATT GTTGTAAAG	4680
TCTAATACCTT CTAAATCCTC AAATGTATTA TCTATTGACG GCTCTAATCT ATTAGTTGTT	4740
AGTGCACCTA AAGATATTT AGATAACCTT CCTCAATTCC TTTCTACTGT TGATTTGCCA	4800
ACTGACCAGA TATTGATTGA GGGTTTGATA TTTGAGGTTC AGCAAGGTGA TGCTTTAGAT	4860
TTTCATTIG CTGCTGGCTC TCAGCGTGGC ACTGTTGCAG GGGGTGTTAA TACTGACCGC	4920
CTCACCTCTG TTTTATCTTC TGCTGGTGGT TCGTTGGTA TTTTTAATGG CGATGTTTTA	4980
GGGCTATCAG TTGCGCATT AAAGACTAAT AGCCATTCAA AAATATTGTC TGTGCCACGT	5040
ATTCTTACGC TTTCAGGTCA GAAGGGTTCT ATCTCTGTTG GCCAGAATGT CCCTTTATT	5100
ACTGGTCGTG TGACTGGTGA ATCTGCCAAT GTAAATAATC CATTTCAGAC GATTGAGCGT	5160
CAAAATGTAG GTATTTCAT GAGCGTTTT CCTGTTGCAA TGGCTGGCGG TAATATTGTT	5220
CTGGATATTA CCAGCAAGGC CGATAGTTIG AGTTCTCTA CTCAGGCAAG TGATGTTATT	5280
ACTAATCAAA GAAGTATTGC TACAACGGTT AATTTCGTG ATGGACAGAC TCTTTACTC	5340
GGTGGCCTCA CTGATTATAA AAACACTTCT CAAGATTCTG GCGTACCGTT CCTGTCTAAA	5400
ATCCCTTAA TCGGCCTCCT GTTTAGCTCC CGCTCTGATT CCAACCGAGGA AAGCACGTTA	5460
TACGTGCTCG TCAAAGCAAC CATAGTA GCGCTGTAGC GGCGCATTAA GCGCGGGGG	5520
TGTGGTGGTT ACGCGCAGCG TGACCGCTAC ACTTGCCAGC GCGCTAGCGC CGCCTCCCTT	5580
CGCTTCTTC CCTTCCTTTC TGGCCACGTT CGCCGGCTTT CCCCGTCAAG CTCTAAATCG	5640
GGGGCTCCCT TTAGGGTTCC GATTTAGTGC TTTACGGCAC CTCGACCCCCA AAAAACTTGA	5700
TTTGGGTGAT GGTCACGTA GTGGGCCATC GCGCTGATAG ACGGTTTTTC GCGCTTGTAC	5760
GTTGGAGTCC ACGTTCTTAA ATAGTGGACT CTGTTCCAA ACTGGAACAA CACTCAACCC	5820
TATCTCGGGC TATTCTTTIG ATTATATAAGG GATTTTGGCG ATTTCGGAAC CACCATCAA	5880
CAGGATTTC GCCTGCTGG GCAAACCAGC GTGGACCGGT TGCTGCAACT CTCTCAGGGC	5940
CAGGGCGTGA AGGGCAATCA GCTGTTGCCG GTCTCCGTGG TGAAAAGAAA AACGACCCCTG	6000
CGGCCCAATA CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCAATTAT CGAGCTGGCA	6060
CGACAGGTTT CCCCAGTGGA AAGCGGGCAG TGAGGCAAC GCAATTATG TGAGTTAGCT	6120
CACTCATTAG GCACCCCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT TGTGTGGAAT	6180
TGTGAGGGGA TAACAATTTC ACACAGGAAA CAGCTATGAC CAGGATGTAC GAATTCCGAG	6240
GTAGGAGAGC TCGGGGGATC CTAGGCTGAA GGGATGACC CTGCTAAGGC TGGATTCAAT	6300
AGTTTACAGG CAAGTGCTAC TGAGTACATT GGCTACGCTT GGGCTATGGT AGTAGTTATA	6360
GTTGGTGCTA CCATAGGGAT TAAATTATTC AAAAAGTTTA CGACCAAGGC TTCTTAACCA	6420
GCTGGCGTAA TAGCGAAGAG GCGCCGACCG ATCGCCCTTC CCAACAGTTG CGCAGCCTGA	6480
ATGGCGAATG GCGCTTGGCC TGGTTCCGG CACCAAGGC GGTGCCGGAA AGCTGGCTGG	6540

AGTGCATCT TCCTGAGGCC GATACGGTCG TCGTCCCCTC AACTGGCAG ATGCACGGTT	6600
ACGATGCGCC CATCTACACC AACGTAACCT ATCCCATTAAC CGTCAATCCG CCGTTTGTTC	6660
CCACGGAGAA TCCGACGGGT TGTTACTCGC TCACATTAA TGTTGATGAA ACCTGGCTAC	6720
AGGAAGGCCA GACGCCAATT ATTTTGATG GCGTTCTAT TGGTAAAAAA ATGAGCTGAT	6780
TTAACAAAAA TTTAACCGA ATTTAACAA AATATTAACG TTTACAATT AAATATTTGC	6840
TTATACAATC TTCTGTTT TGGGGCTTT CTGATTATCA ACCGGGGTAC ATATGATTGA	6900
CATGCTAGTT TTACGATTAC CGTTCATCGA TTCTCTTGT TGCTCCAGAC TCTCAGGCCA	6960
TGACCTGATA GCCTTTGTAG ATCTCTAAA AATAGCTACC CTCTCCGGCA TTAATTATC	7020
AGCTAGAACG GTTGAATATC ATATTGATGG TGATTGACT GTCTCCGGCC TTTCTCACCC	7080
TTTTGAATCT TTACCTACAC ATTACTCAGG CATTGCATT AAAATATATG AGGGTTCTAA	7140
AAATTTTAT CCTTGCCTTG AAATAAAGGC TTCTCCGGCA AAAGTATTAC AGGGTCATAA	7200
TGTTTTGGT ACAACCGATT TAGCTTATG CTCTGAGGCT TTATTGCTTA ATTTGCTAA	7260
TTCTTGCCT TGCCTGTATG ATTTATTGGA CGTT	7294

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7320 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTTCAG CTCGGCCCC AAATGAAAAT	60
ATAGCTAAC AGGTTATTGA CCATTGCGA AATGTATCTA ATGGTAAAC TAAATCTACT	120
CGTTCCGAGA ATTGGGAATC AACTGTTACA TGGAAATGAAA CTTCCAGACA CCGTACTTTA	180
GTTGCATATT TAAAACATGT TGAGCTACAG CACCAAGATTG AGCAATTAAG CTCTAAGCCA	240
TCTGAAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG	300
TGGAGTTTG CTTCCGGTCT GGTTGCTTT GAAGCTCGAA TTAAAACGCG ATATTGAAAG	360
TCTTCCGGGC TTCTCTTAA TCTTTTGAT GCAATCCGCT TTGCTCTGA CTATAATAGT	420
CAGGGTAAAG ACCTGATTT TGATTATGG TCATTCTCGT TTTCTGAACG GTTAAAGCA	480
TTTGAGGGGG ATTCAATGAA TATTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT	540
AAACATTTA CTATTACCCC CTCTGGAAA ACTTCTTTG CAAAGCCTC TCGTATTTT	600
GGTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC TATGCCTCGT	660
AATTCTTTT GCGCTTATGT ATCTGCATTA GTTGAATGTG GTATTCTAA ATCTCAACTG	720
ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA CGTAGATTTT	780
TCTTCCCAAC GTCCTGACTG GTATAATGAG CCAGTTCTTA AAATCCGATA AGGTAATTCA	840

CAATGATTAA AGTTGAAATT AA	CCATOTC AAGGCCAATT TACTACTCGT	TCTGGTGT	900	
CTCGTCAGGG CAAGCCTTAT	TCACTGAATG AGCAGCTTG TTACGTTGAT	TTGGGTAATG	960	
AATATCCGGT	TCTTGTCAAG ATTACTCTTG ATGAAGGTCA	GCCAGCCTAT	1020	
TGTACACCGT	TCATCTGTCC TCTTCAAAG TTGGTCAGTT	CGGTTCCCTT	1080	
GTCTGCCCT CGTTCCGGCT	AAGTAACATG GAGCAGGTGG CGGATTTCGA	CACAATTAT	1140	
CAGGCGATGA	TACAAATCTC CGTTGTACTT TGTTTCCGCC	TTGGTATAAT	1200	
CAAAGATGAG	TGTTTAGTG TATTCTTCG CCTCTTCG	TTAGGTTGG	1260	
GTGGCATTAC	GTATTTACC CGTTAAATGG	AAACTTCCTC	1320	
CAAAGCCTCT GTAGCCGTTG	CTACCCCTCGT	TCCGATGCTG	1380	
CGATCCCCGA	AAAGCGGCCT	TTAACTCCCT	CGAACCCCTCA	1440
TGGCTGGCG ATGGTTGTTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	1500
ATTACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	1560
TTTTGGAGA	TTTCAACGT	GAAAAAATTA	TTATTCGAA	1620
TATTCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTAGCAA	1680
TTTACTAACG	TCTGGAAAGA	CCACAAA	TTAGATCGTT	1740
CTGTGGAATG	CTACAGCGT	TGTAGTTGT	ACTCGTGACG	1800
TGGGTTCTA	TTGGGTTG	TATCCCTGAA	AATGAGGGTG	1860
TCTGAGGGTG	CGGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	1920
ATTCCGGGCT	ATACTTATAT	CAACCGCTC	GACGGCACTT	1980
AACCCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	2040
CAGAATAATA	GGTTCCGAAA	TAGGCAGGGG	GCATTAACG	2100
CAAGGCAGTG	ACCCCGTTAA	AACTTATTAC	CACTACACTC	2160
TATGACGGTT	ACTGGAACGG	TAAATTCTAGA	GAUTGGCCTT	2220
GATCCATTCC	TTTGTGAATA	TCAAGGCCAA	TCGTCTGACC	2280
GCTGGGGCG	GCTCTGGTGG	TGGTCTG	GGCGGCTCTG	2340
GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGA	GGCGGTTCCG	2400
GATTTGATT	ATGAAAAGAT	GGCAAACGCT	AAATAAGGGGG	2460
GAAAACGCC	TACAGTCTGA	CCCTAAAGGC	AAACTTGATT	2520
GCTGCTATCG	ATGGTTTCAT	TGGTGACGTT	TCCGGCCTTG	2580
GGTGATTTG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCG	2640
TTAATGAATA	ATTTCCGTCA	ATATTTACCT	TCCCTCCCTC	2700
TTTGTCTTAA	GGGCTGGTAA	ACCATATGAA	TTTCTATTG	2760
TTCCGTGGTG	TCTTTCGTT	TCTTTTATAT	GTTGCCACCT	2820
TTTGCTAACAA	TACTGCGTAA	TAAGGAGTCT	TAATCATGCC	2880

TATTATTGCG TTTCCCTCGGT TTCCCTCTGG TAACTTTGTT CGGCTATCTG CTTACTTTTC	2940
TTAAAAAGGG CTTCGGTAAG ATAGCTATTG CTATTCATT GTTCTTGCT CTTATTATTG	3000
GGCTTAACTC AATTCTTGTG GGTTATCTCT CTGATATTAG CGCTCAATTAA CCCTCTGACT	3060
TTGTTCAAGGG TGTTCAAGTTA ATTCTCCCGT CTAATGCGCT TCCCTGTTTT TATGTTATTG	3120
TCTCTGTAAA GGCTGCTATT TICATTGTT AGGTTAAACA AAAAATCGTT TCTTATTG	3180
ATTGGGATAA ATAATATGGC TGTTTATTGTT GTAACTGGCA AATTAGGCTC TCGAAAGACG	3240
CTCGTTAGCG TTGGTAAGAT TTAGGATAAA ATTGTAGCTG CGTGCAAAAT ACCAACTAAT	3300
CTTGATTTAA GGCTTCAAAA CCTCCCGCAA GTCGGGAGGT TCGCTAAAC GCCTCGCGTT	3360
CTTAGAATAC CGGATAAGCC TTCTATATCT GATTGCTTG CTATTGGGG CGGTAAATGAT	3420
TCCTACGATG AAAATAAAAA CGGCTTGCTT GTTCTCGATG AGTGGGTAC TTGGTTTAAT	3480
ACCCGTTCTT GGAATGATAA GGAAAGACAG CCGATTATTG ATTGGTTCT ACATGCTCGT	3540
AAATTAGGAT GGGATATTAT CTTCTTGTGTT CAGGACTTAT CTATTGTTGA TAAACAGGCG	3600
CGTTCTGCAT TAGCTGAACA TGTTGTTTAT TGTCGTGTC TGGACAGAAAT TACTTTACCT	3660
TTTGTCCGTA CTTTATATTC TCTTATTACT GCCTCGAAAA TGCCTCTGCC TAAATTACAT	3720
GTGGCGTTG TTAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGCG TTGGTTTAT	3780
ACTGGTAAGA ATTGTATAA CGCATATGAT ACTAAACAGG CTTTTCTAG TAATTATGAT	3840
TCCGGTGTGTT ATTCTTATTGTT AACGCCTTAT TTATCACACG GTCGGTATT CAAACCATT	3900
AATTAGGTC AGAAGATGAA ATTAACTAAA ATATATTGAA AAAAGTTTTC TCGCGTTCTT	3960
TGTCTTGCAGA TTGGATTGTC ATCAGCATT ACATATAGTT ATATAACCCA ACCTAAGCCG	4020
GAGGTTAAAA AGGTAGTCTC TCAGACCTAT GATTGATA AATTCACTAT TGACTCTTCT	4080
CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAAAT	4140
AGCGACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTATG TACTGTTCC	4200
ATTAAAAAAG GTAATTCAA TGAAATTGTT AAATGTAATT AATTGTTTT TCTTGATGTT	4260
TGTTTCATCA TCTCTTTG CTCAGGTAAT TGAAATGAAT AATTGCGCTC TGCGCGATT	4320
TGTAACCTGG TATTCAAAGC AATCAGGGGA ATCCGTTATT GTTCTCCCG ATGTAAGG	4380
TACTGTTACT GTATATTCACT CGACGTTAA ACCTGAAAAT CTACGCAATT TCTTATTTC	4440
TGTTTACGT GCTAATAATT TTGATATGGT TGTTCAATT CCTTCCATAA TTCAGAAAGTA	4500
TAATCCAAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATGA	4560
TGATAATTCC GCTCCTCTG GTGGTTCTT TGTTCCGAA AATGATAATG TTACTCAAAC	4620
TTTTAAAATT AATAACGTTG GGGCAAAGGA TTTAATACGA GTTGTGAAAT TGTTGTAAA	4680
GTCTAATACT TCTAAATCCT CAAATGTATT ATCTATTGAC GGCTCTAAC TATTAGTTGT	4740
TAGTGCACCT AAAGATATTG TAGATAACCT TCCTCAATTG CTTTCTACTG TTGATTG	4800
AACTGACCAAG ATATTGATTG AGGGTTGAT ATTTGAGGTT CAGCAAGGTG ATGCTT	4860
TTTTTCATTT GCTGCTGGCT CTCAGCGTGG CACTGTTGCA GGCGGTGTTA ATACTGACCG	4920

CCTCACCTCT	GT	TTTATCTT	CTGCTCGTGG	TTCGTTGGT	ATTTTAATG	GGCATGTTT	4980
AGGGCTATCA	GT	TTCGCGCAT	TAAAGACTAA	TAGCCATTCA	AAAATATTGT	CTGTGCCACG	5040
TATTCTTACG	CT	TTTCAGGTC	AGAAGGGTTC	TATCTCTGT	GGCCAGAATG	TCCCTTTAT	5100
TACTGGTCGT	GT	GACTGGTG	AATCTGCCAA	TGTAAATAAT	CCATTTCAGA	CGATTGAGCG	5160
TCAAAATGTA	GGT	ATTTCCA	TGAGCGTTT	TCCTGTTGCA	ATGGCTGGCG	GTAATATTGT	5220
TCTGGATATT	ACCAGCAAGG	CCGATAGTTT	GAGTTCTTCT	ACTCAAGCAA	GTGATGTTAT	5280	
TACTAATCAA	ACAAGTATTG	CTACAACGGT	TAATTTGCGT	GATGGACAGA	CTCTTTACT	5340	
CGGTGGCCTC	ACTGATTATA	AAAACACTTC	TCAAGATTCT	GGCGTACCGT	TCCTGTCTAA	5400	
AATCCCTTTA	ATCGGGCTCC	TGTTAGCTC	CCGCTCTGAT	TCCAACGAGG	AAAGCACGTT	5460	
ATACGTGCTC	GTCAAAGCAA	CCATAGTACG	CGCCCTGTAC	GGGCCATTAA	AGCCGGGGGG	5520	
GTGTGGTGGT	TACGCCAGC	GTGACCGCTA	CACTTGCAG	GGCCCTAGGG	CCCGCTCCTT	5580	
TCGCTTCTT	CCCTTCCTT	CTCGCCACGT	TCGCCCCCTT	TCCCCGTCAA	GCTCTAAATC	5640	
GGGGGCTCCC	TTTAGGGTTC	CGATTAGTG	CTTTACGGCA	CCTCGACCCC	AAAAAAACTTG	5700	
ATTTGGTGA	TGGTCACGT	ACTGGCCAT	CGCCCTGATA	GACGTTTTT	GGCCCTTGA	5760	
CGTTGGAGTC	CACGTTCTT	AATAGTGGAC	TCTTGTCCA	AACTGGAACA	ACACTCAACC	5820	
CTATCTCGGG	CTATTCTTT	GATTTATAAG	GGATTTTGCC	GATTTGGAA	CCACCATCAA	5880	
ACAGGATTT	CGCCTGCTGG	GGCAAACACAC	GTGGACCGC	TTGCTGCAAC	TCTCTCAGGG	5940	
CCAGGGGGTG	AAGGGCAATC	ACCTTTGCC	CGTCTCGCTG	GTGAAAAGAA	AAACCACCC	6000	
GGCGCCCAAT	ACGCAAACCG	CCTCTCCCCG	CGCGTGGGCC	GATTCAATTAA	TGCAGCTGGC	6060	
ACGACAGGTT	TCCCGACTGG	AAAGCGGGCA	GTGAGCGCAA	CGCAATTAAAT	GTGAGTTAGC	6120	
TCACTCATTAA	GGCACCCCCAG	GTCTTACACT	TTATGTTCC	GGCTCGTATG	TTGTGTGGAA	6180	
TTGTGAGCGG	ATAACAATTI	CACACGCCAA	GGAGACAGTC	ATAATGAAAT	ACCTATTGCC	6240	
TACGGCAGCC	GCTGGATTGT	TATTACTCGC	TGCCCAACCA	GCCATGGGG	AGCTCGTGAT	6300	
GACCCAGACT	CCAGAATTCC	ATCCGAATG	AGTGTAAATT	CTAGAACCGC	TAAGCTTGGC	6360	
ACTGGCCGTC	GT	TTTACAAAC	GTCTGACTG	GGAAAACCT	GGCGTTACCC	AACTTAATCG	6420
CCTTGCAGCA	CACCCCCCTT	TCGCCAGCTG	GGCTAATAGC	GAAGAGGCC	GCACCGATCG	6480	
CCCTTCCCAA	CAGTTGGC	GA	CGTGAATGG	CGAATGGGC	TTTGCCTGGT	TCGGCACC	6540
AGAAGCGGTG	CCGGAAAGCT	GGCTGGAGTC	CGATCTTCT	GAGGCCGATA	GGGTGCGTGT	6600	
CCCTCTAAAC	TGGCAGATGC	ACGGTTACGA	TGCGCCCAC	TACACCAACG	TAACCTATCC	6660	
CATTACGGTC	AATCCGCCGT	TTGTTCCCAC	GGAGAATCCG	ACGGGTTGTT	ACTCGCTCAC	6720	
ATTTAATGTT	GATGAAAGCT	GGCTACAGGA	AGGCCAGACG	CGAATTATTT	TTGATGGCGT	6780	
TCCTATTGGT	AAAAAAATCA	GCTGATTAA	CAAAAATTAA	ACGCGAATT	TAACAAAATA	6840	
TTAACGTTTA	CAATTAAAT	ATTTGCTTAT	ACAATCTCC	TGTTTTGGG	GCTTTCTGA	6900	
TTATCAACCG	GGGTACATAT	GATTGACATG	CTAGTTTAC	GATTACCGTT	CATCGATTCT	6960	

CTTGTGCT CCAGACTCTC AGGCAATGAC CTGATAGCCT TTGTAGATCT CTCAAAAAATA	7020
GCTACCCCTCT CCGGCATTAAT TTTATCAGCT AGAACGGTTG AATATCATAT TGATGGTGT	7080
TTGACTGTCT CCGGCCTTTC TCACCCCTTT GAATCTTAC CTACACATTA CTCAGGCATT	7140
GCATTTAAAA TATATGAGGG TTCTAAAAAT TTTTATCCTT GCGTCAAAT AAAGGCTTCT	7200
CCCGCAAAAG TATTACAGGG TCATAATGTT TTGCTACAA CCGATTAGC TTTATGCTCT	7260
GAGGCTTAT TGCTTAATTCT TTGCTTGCC TGTATGATTT ATTGGACGTT	7320

(2) INFORMATION FOR SEQ ID NO:3:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7445 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTCAAG CTCGCGCCCC AAATGAAAAT	60
ATAGCTAAC AGGTATTGAA CCATTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT	120
CGTTCCGAGA ATTGGGAATC AACTGTTACA TCGAATGAAA CTTCCAGACA CCGTACTTTA	180
GTTGCATATT TAAAACATGT TGAGCTACAG CACCAGATTC AGCAATTAAG CTCTAAGCCA	240
TCTGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG	300
TTGGAGTTTG CTTCCGGTCT GGTTCCGTTT GAAGCTCGAA TTAAAACGCG ATATTGAAAG	360
TCTTTCGGGC TTCCCTCTTAA TCTTTTGAT GCAATCCGCT TTGCTCTGA CTATAATAGT	420
CAGGGTAAAG ACCTGATTT TGATTATGG TCATTCTCGT TTTCTGAACG GTTAAAGCA	480
TTTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT	540
AAACATTTTA CTATTACCCC CTCTGGCAAA ACTTCTTTG CAAAAGCCTC TCGCTATTCT	600
GGTTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC TATGCCCTCGT	660
AATTCTTTT GCGCTTATGT ATCTGCATTA GTTGAATGTG GTATTCTAA ATCTCAACTG	720
ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTG GTTTTATTAA CGTAGATTTT	780
TCTTCCAAC GTCCGTACTG GTATAATGAG CCGAGTTCTTA AAATCCGATA AGGTAATTCA	840
CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCCAATT TACTACTCGT TCTGGTCTT	900
CTCGTCAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTGT TTACGTTGAT TTGGGTAATG	960
AATATCCGGT TCTTGTCAAG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT GCGCCTGGTC	1020
TGTACACCGT TCATCTGTCC TCTTCAAAG TTGGTCAGTT CGGTTCCCTT ATGATTGACC	1080
GTCTGCCCT CGTTCCGGCT AAGTAACATG GAGCAGGTGG CGGATTTCGA CACAATTAT	1140
CAGGGCATGA TACAAATCTC CGTTGTACTT TGTTTCCGCG TTGGTATAAT CGCTGGGGGT	1200
CAAAGATGAG TGTTTAGTG TATTCTTCG CCTCTTCTCGT TTTAGGTTGG TGCCTTCGTA	1260

GTGGCATTAC GTATTTTACCGTTTAAATGG	1320
AAACTTCGCTG ATGAAAAAGT CTTTAGTCCT	
CAAAGCCTCT GTAGCCGTTG CTACCGCTGT TCCGATGCTG	1380
TCTTTCGCTG CTGAGGGTGA	
CGATCCCGCA AAAGCGGCCT TTAACTCCCT GCAAGCCTCA	1440
GGCACCGAAT ATATCGTTA	
TGGGTGGCG ATGGTTGTTG TCATTGTCGG CGCAACTATC	1500
GGTATCAAGC TGTAAAGAA	
ATTCACCTCG AAAGCAAGCT GATAAACCGA TACAATTAAA	1560
GGCTCCTTT GGAGCCTTT	
TTTTGGAGA TTTTCAACGT GAAAAAATTA TTATTCGCAA	1620
TTCCTTACTG TGTAAAGT TGTTAGCAA AACCCCATAC	
AGAAAATTCA	1680
TATTCTCACT CCGCTGAAAC TGTTGAAAGT TGTTAGCAA	
AACCCCATAC AGAAAATTCA	1740
TTTACTAACG TCTGGAAAGA CGACAAAAGT TTAGATCGTT	
ACGCTAACTA TGAGGGTTGT	1800
CTGTGGAATG CTACAGGGT TGTACTTGT ACTGGTGAGG	
AAACTCAGTG TTACGGTACA	1860
TGGGTTCTA TTGGGCTTGC TATCCCTGAA AATGAGGGTG	
GTGGCTCTGA GGGTGGCGGT	1920
TCTGAGGGTG GCGGTTCTGA GGGTGGCGGT ACTAAACCTC	
CTGAGTACGG TGATACACCT	1980
ATTCCGGGCT ATACTTATAT CAACCCCTCTC GACGGCACTT	
ATCCGCCTGG TACTGAGCAA	2040
AAACCCGCTA ATCCTAATCC TTCTCTTGAG GAGTCTCAGG	
CTCTTAATAC TTTCATGTTT	2100
CAGAATAATA GGTTCCGAAA TAGGCAGGGG GCATTAACGT	
TTTATACGGG CACTGTTACT	2160
CAAGGCAGTG ACCCCGTTAA AACTTATTAC CAGTACACTC	
CTGTATCATC AAAAGCCATG	2220
TATGACCGCTT ACTGGAACGG TAAATTCTAGA GACTGGCTT	
TCCATTCTGG CTTTAATGAA	2280
GATCCATTCTG TTTGTGAATA TCAAGGCCAA TCGTCTGACC	
TGCCTCAACC TCCGTCAAT	2340
GCTGGCGCG GCTCTGGTGG TGTTCTGGT GGCGGCTCTG	
AGGGTGGTGG CTCTGAGGGT	2400
GGCGGTTCTG AGGGTGGCGG CTCTGAGGGA GGGGTTCCG	
GTGGTGGCTC TGGTCCGGT	2460
GATTTTGATT ATGAAAAGAT GGCAACCGCT AATAAGGGGG	
CTATGACCGA AAATGCCGAT	2520
AAAAACGGCG TACAGTCTGA CGCTAAAGGC AAACCTGATT	
CTGTCGCTAC TGATTACGGT	2580
GCTGCTATCG ATGGTTTCAT TGGTGACGTT TCCGGCTTG	
CTAATGGTAA TGGTGTACT	2640
GGTGATTTG CTGGCTCTAA TTCCCAAATG GCTCAAGTCG	
GTGACGGTGA TAATTACACCT	2700
TTAATGAATA ATTTCCGTCA ATATTCACCT TCCCTCCCTC	
AATCGGTGA ATGTCGCCCT	2760
TTTGTCTTAA GCGCTGGTAA ACCATATGAA TTTCTATTG	
ATTGTGACAA AATAAAACTTA	2820
TTCCGTGGTG TCCTTGCGTT TCTTTATAT GTTGGCACCT	
TTATGTATGT ATTTCTACG	2880
TTTGCCTAACA TACTGGTAA TAAGGAGTCT TAATCATGCC	
AGTTCTTTG GGTATTCCGT	2940
TATTATTGCG TTTCCCTGGT TTCCCTCTGG TAACCTTGTG	
CGGCTATCTG CTTACTTTTC	3000
TTAAAAAGGG CTTCGGTAAAG ATAGCTATTG CTATTCATT	
GTTCCTTGCT CTTATTATTG	3060
GGCTTAACTC AATTCTTGTG CGTTATCTCT CTGATATTAG	
CGCTCAATTAA CCCTCTGACT	3120
TTGTTCAAGGG TGTTCAAGTTA ATTCTCCGT CTAATGGCT	
TCCCTGTTTT TATGTTATTG	3180
TCTCTGTAAA GGCTGCTATT TTCATTTTG ACGTTAAACA	
AAAAATCGTT TCTTATTG	3240
ATTGGGATAAA ATAATATGGC TGTAAAGAT TCAGGATAAA	
ATTGTAGCTG GGTGAAAGACG	3300
CTCGTTAGCG TTGGTAAGAT TCAGGATAAA ATTGTAGCTG	
GGTGCAGGAAAT AGCAACTAAT	

CTTGATTTAA GGCTTCAAAA CCTCCCGCAA GTCGGGAGGT TCGCTAAAAC GCCTCGCGTT	3360
CTTAGAATAC CGGATAAGCC TTCTATATCT GATTGCTTG CTATTGGCG CGCTAATGAT	3420
TCCTACGATG AAAATAAAAA CGGCTTGCTT GTTCTCGATG AGTGCCTAC TTGGTTTAAT	3480
ACCCGTTCTT GGAATGATAA GCAAAGACAG CCGATTATTG ATTGGTTCT ACATGCTCGT	3540
AAATTAGGAT GGGATATTAT TTTTCTTGT CAGGACTTAT CTATTGTTGA TAAACAGGCC	3600
CGTTCTGCAT TAGCTGAACA TGTGTTAT TGTCTCGTC TGGACAGAAAT TACTTACCT	3660
TTTGTGGTA CTTTATATTCT CTTTATTACT GGCTCGAAAA TGCCCTGCC TAAATTACAT	3720
GTTGGCGTTG TAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGCG TTGGCTTAT	3780
ACTGGTAAGA ATTTGTATAA CGCATATGAT ACTAAACAGG CTTTTCTAG TAATTATGAT	3840
TCCGGTGT ATTCTTATTAAACGCCCTAT TTATCACACG GTCGGTATT CAAACCATTAA	3900
AATTAGGTC AGAAGATGAA GCTTACTAAA ATATATTGAA AAAAGTTTC ACGGCTTCTT	3960
TGTCTTCCGA TTGGATTTCG ATCAGCATT ACATATACTT ATATAACCCA ACCTAACCG	4020
GAGGTTAAAA AGGTAGTCTC TCAGACCTAT GATTTGATA AATTCACTAT TGACTCTCT	4080
CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAA	4140
AGCGACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTTATG TACTGTTCC	4200
ATTAAAAAG GTAATTCAA TGAAATTGTT AAATGTAATT AATTTGTTT TCTTGATGTT	4260
TGTTTCATCA TCTCTTTTG CTCAGGTAAT TGAAATGAAT AATTCGCTC TGCGCGATT	4320
TGTAACCTGG TATTCAAAGC AATCAGGCCA ATCCGTTATT GTTCTCCCG ATGAAAAGG	4380
TACTGTTACT GTATATTCACTGACGTTAA ACCTGAAAAT CTACGCAATT TCTTATTTC	4440
TGTTTACGT GCTAATAATT TTGATATGGT TGGTCAATT CCTTCATAA TTCAAGAGTA	4500
TAATCCAAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATGA	4560
TGATAATTCC GCTCCTCTG GTGGTTCTT TGTCCGCAA AATGATAATG TTACTCAAAC	4620
TTTTAAAATT AATAACGTTG GGGCAAAGGA TTTAATACGA GTTGTGAAT TGTTGTAAA	4680
GTCTAAACTCT CAAATGTTTATCCTT ATCTATTGAC GGCTCTAAC TATTAGTTGT	4740
TAGTGACCT AAAGATATT TAGATAACCT TCCCTCAATTCTTCTACTG TTGATTGCC	4800
AACTGACCAAG ATATTGATTG AGGGTTGAT ATTTGAGGTT CAGGAAGGTG ATGTTTACA	4860
TTTTGATTT GCTGCTGGCT CTCAGCGTGG CACTGTTGCA GGCGGTGTTA ATACTGACCG	4920
CCTCACCTCT GTTTTATCTT CTGCTGGTGG TTGCTCGGT ATTTTAATG GCGATGTTT	4980
AGGGCTATCA GTTCGCGCAT TAAAGACTAA TAGCCATTCA AAAATATTGT CTGTGCCACG	5040
TATTCTTACG CTTTCAGGTC AGAAGGGTTC TATCTCTGTT GGCCAGAATG TCCCTTTAT	5100
TACTGGTCGT GTGACTGGTG AATCTGCCAA TGAAATAAT CCATTTCAGA CGATTGAGCG	5160
TCAAAATGTA GGTATTCCA TGAGCGTTT TCCGTGCA ATGGCTGGCG GTAATATTGT	5220
TCTGGATATT ACCAGCAAGG CCGATAGTTT GAGTTCTCT ACTCAGGCAA GTGATGTTAT	5280
TACTAATCAA AGAAAGTATTG CTACAACGGT TAATTGCGT GATGGACAGA CTCTTTACT	5340

CGGTGGCCTC ACTGATTATA AAAACACTTC TCAAGATTCT GGCgtaccgt TCCTGTCTAA 5400
 AATCCCTTTA ATCGGCCTCC TGTTTAGCTC CCGCTCTGAT TCCAACGAGG AAAGCACGTT 5460
 ATACGTGCTC GTCAAAGCAA CCATACTACG CGCCCTGTAG CGGCCATTAA AGCGCGGGGG 5520
 GTGTGGTGGT TACGCGCAGC GIGACGGCTA CACTTGCCAG CGCCCTAGGG CCCGCTCCTT 5580
 TCGCTTCTT CCCTTCCTT CTCGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC 5640
 GGGGGCTCCC TTTAGGGTTC CGATTAGTC CTTTACGGCA CCTCGACCCC AAAAAACTTG 5700
 ATTTGGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTTGA 5760
 CGTTGGAGTC CACGTTCTT AATAGTGGAC TCTTGTCTCA AACTGGAACA ACACCTCAACC 5820
 CTATCTCGGG CTATTCTTT GATTATAAG GGATTTGCC GATTTGGAA CCACCATCAA 5880
 ACAGGATTTT CGCCTGCTGG GGCAAACCAAG CGTGGACCCG TTGCTGCAAC TCTCTCAGGG 5940
 CCAGGGGGTG AAGGGCAATC AGCTGTTGCC CGTCTCGCTC GTGAAAAGAA AAACCACCC 6000
 GCCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGCC GATTCAATTAA TGCAGCTGGC 6060
 ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGGCCA CGCAATTAAT GTGAGTTAGC 6120
 TCACTCAATTAA GGCACCCCCAG GCTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA 6180
 TTGTGAGCGG ATAACAATT T CACACGCGTC ACTTGGCACT GGCCGTCGTT TTACAACGTC 6240
 GTGACTGGGA AAACCCCTGGC GTTACCCAAG CTTTGTACAT GGAGAAAATA AAGTGAAACA 6300
 AAGCACTATT GCACTGGCAC TCTTACCGTT ACCGTTACTG TTTACCCCTG TGACAAAAGC 6360
 CGCCCAAGGTC CAGCTGCTCG AGTCAGGCCT ATTGTGCCA GGGGATTGTA CTAGTGGATC 6420
 CTAGGCTGAA GGCGATGACC CTGCTAAGGC TGCATTCAAT AGTTTACAGG CAAGTGCTAC 6480
 TGAGTACATT GGCTACGCTT GGGCTATGGT ACTAGTTATA GTGGTGCTA CCATAGGGAT 6540
 TAAATTATTC AAAAAGTTA CGAGCAAGGC TTCTTAAGCA ATAGCGAAGA GGCCCCGACC 6600
 GATGCCCTT CCCAACAGTT GCGCAGCCTG AATGGCGAAT GGCGCTTGC CTGGTTCCG 6660
 GCACCGAGAAG CGGTGCCGGGA AAGCTGGCTG GAGTGGCATC TTCTGAGGC CGATACGGTC 6720
 GTCGTCCCT CAAACTGGCA GATGCACGGT TACGATGCCG CCATCTACAC CAACGTAACC 6780
 TATCCCATTA CGGTCAATCC GCCGTTTGT CCCACGGAGA ATCCGACGGG TTGTTACTCG 6840
 CTCACATTAA ATGTTGATGA AAGCTGGCTA CAGGAAGGCC AGACGCGAAT TATTTTGAT 6900
 GGCCTTCTA TTGGTTAAAA AATGAGCTGA TTTAACAAAA ATTTAACGCG AATTTTAACA 6960
 AAATATTAAC GTTTACAATT TAAATATTG CTTATACAAT CTTCTGTCTT TTGGGGCTTT 7020
 TCTGATTATC AACCGGGCTA CATATGATTG ACATGCTAGT TTTACGATTA CGGTTCATCG 7080
 ATTCTCTTGT TTGCTCCAGA CTCTCAGGCA ATGACCTGAT AGCCTTCTA GATCTCTCAA 7140
 AAATAGCTAC CCTCTCCGGC ATTAATTAT CAGCTAGAAC GGTTGAATAT CATATTGATG 7200
 GTGATTGAC TGTCTCCGGC CTTTCTCACC CTTTGAATC TTTACCTACA CATTACTCAG 7260
 GCATTGCATT TAAAATATAT GAGGGTTCTA AAAATTCTA TCCTTGCGTT GAAATAAAGG 7320
 CTTCTCCGGC AAAAGTATTAA CAGGGTCATA ATGTTTIGG TACAACCGAT TTAGCTTTAT 7380

GCTCTGAGGC TTTATTGCTT AATTTGCTA ATTCTTGCC TTGCCTGTAT GATTTATTGG	7440
ACGTT	7445

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 7409 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTCAAG CTCGGCCCC AAAAAGAAAAT	60
ATAGCTAAAC AGGTTATTGA CCATTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT	120
CGTTCCAGA ATTGGGAATC AACTGTTACA TGGAATGAAA CTTCCAGACA CCGTACTTTA	180
GTTGCATATT TAAAACATGT TGAGCTACAG CACCAGATTC AGCAATTAAG CTCTAAGCCA	240
TCTGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG	300
TTGGAGTTTG CTTCGGTCT GGTTGGCTTT GAAGCTCGAA TTAAAACGCG ATATTTGAAG	360
TCTTTGGGC TTCCCTCTAA TCTTTTGAT GCAATCCGCT TTGCTCTGA CTATAATAGT	420
CAGGGTAAAG ACCTGATTT TGATTTATGG TCATTCTCGT TTTCTGAACG GTTTAAAGCA	480
TTTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGAG TATTGGACGC TATCCAGTCT	540
AAACATTTTA CTATTACCCC CTCTGGAAA ACTTCTTTG CAAAAGCCTC TCGCTATTTT	600
GGTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTAC TATGCTCGT	660
AATTCTTTT GGCGTTATGT ATCTGCATTA GTTGAATGTG GTATTCTAA ATCTCAACTG	720
ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA CGTAGATTIT	780
TCTTCCCAAC GTCTGACTG GTATAATGAG CCAGTTCTTA AAATCCGATA AGGTAAATTCA	840
CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGT	900
CTCGTCAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTG TTACGTTGAT TTGGGTAAATG	960
AATATCCGGT TCTTGTCAAG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT GCGCCTGGTC	1020
TGTACACCGT TCATCTGTCC TCTTCAAAG TTGGTCAGTT CGGTTCCCTT ATGATTGACC	1080
GTCTGGCCT CGTTCCGGCT AAGTAACATG GAGCAGGTGG CGGATTCGA CACAATTAT	1140
CAGGCGATGA TACAAATCTC CGTTGTACTT TGTTGGCC TTGGTATAAT CGCTGGGGGT	1200
CAAAGATGAG TGTTTAGTG TATTCTTCG CCTCTTTCGT TTTAGGTGG TGCCTTCGTA	1260
GTGGCATTAC GTATTTTAC CGTTTAATGG AAACCTCCTC ATGAAAAAGT CTTAGTCCT	1320
CAAAGCCTCT GTAGCCGTTG CTACCCCTCGT TCCGATGCTG TCTTGGCTG CTGAGGGTGA	1380
CGATCCCCCA AAAGCGGCCT TTAACCCCT GCAAGCCTCA GCGACCGAAT ATATCGGTTA	1440
TGGCTGGGGG ATGGTTGTTG TCATTGTCGG CCCAACTATC GGTATCAAGC TGTTAAAGAA	1500

ATTCACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	GGCTCCTTTT	CGAGCCTTTT	1560
TTTTGGAGA	TTTCAACGT	GAAAAAATTA	TTATTCGAA	TTCTTACT	TGTTCCCTTC	1620
TATTCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTTAGCAA	AACCCCATAC	AGAAAATTCA	1680
TTTACTAACG	TCTGGAAAGA	CGACAAAAC	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT	1740
CTGTGGAATG	CTACAGGCCT	TGTAGTTGT	ACTGGTGAGC	AAACTCAGTG	TTACGGTACA	1800
TGGGTTCTA	TTGGGCTTGC	TATCCGTGAA	AATGAGGGTC	GTGGCTCTGA	GGGTGGCGGT	1860
TCTGAGGGTG	CGCGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGACTACGG	TGATACACCT	1920
ATTCCGGCT	ATACTTATAT	CAACCCCTCTC	GACGGCACTT	ATCCGGCTGG	TACTGAGCAA	1980
AACCCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCATGTTT	2040
CAGAATAATA	GGTTCCGAAA	TAGGCAGGGG	GCATTAACTG	TTTATACGGG	CACTGTTACT	2100
CAAGCCACTG	AGCCCGTTAA	AACTTATTAC	CACTACACTC	CTGTATCATC	AAAAGCCATG	2160
TATGACGCTT	ACTGGAACGG	TAAATTAGA	GAATGCGCTT	TCCATTCTGG	CTTTAATGAA	2220
GATCCATTGG	TTTGTGAATA	TCAAGGCCAA	TCGTCTGACC	TGCCTCAACC	TCCTGTCAAT	2280
GCTGGCGGGC	GCTCTGGTGG	TGGTTCTGGT	GGCGGCTCTC	AGGGTGGTGG	CTCTGAGGGT	2340
GGCGGTTCTG	AGGGTGGCCCG	CTCTGAGGGA	GGCGGTTGG	GTGGTGGCTC	TGTTCCGGT	2400
GATTTGATT	ATGAAAAGAT	GGCAAACGCT	ATAAGGGGG	CTATGACCGA	AAATGCCGAT	2460
GAAAACGCCG	TACAGTCTGA	CGCTAAAGGC	AAACTTGATT	CTGTCGCTAC	TGATTACGGT	2520
GCTGCTATCG	ATGGTTTCAT	TGGTGACGTT	TCCGGCCTTG	CTAATGCTAA	TGGTGCTACT	2580
GGTGATTTG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTGC	GTGACGGTGA	TAATTCACCT	2640
TTAATGAATA	ATTTGGTCA	ATATTACCT	TCCCTCCCTC	AATCGGTGA	ATGTCGCCCT	2700
TTTGTCTTAA	GCGCTGGTAA	ACCATATGAA	TTTCTATTG	ATTGTGACAA	AATAAAACTTA	2760
TTCCGTGGTG	TCTTTGGGTT	TCTTTATAT	CTTGCCACCT	TTATGTATGT	ATTTCTACCG	2820
TATTATTGCC	TTCCCTCGGT	TTCCCTCTGG	TAACTTTGTG	CGGCTATCTG	CTTACTTTTC	2880
TTAAAAAGGG	CTTCGGTAAG	ATAGCTATTG	CTATTCATT	GTTCCTTGCT	CTTATTATTG	2940
GGCTTAACTC	AATTCTTGTG	GGTTATCTCT	CTGATATTAG	CGCTCAATTAA	CCCTCTGACT	3000
TTGTCAGGG	TGTCAGTTA	ATTCCTCCGT	CTAATGCCCT	TCCCTGTTT	TATGTTATTC	3060
TCTCTGTAAA	GGCTGCTATT	TTCACTTTTG	ACGTTAAACA	AAAAATCGTT	TCTTATTTGG	3120
ATTGGCATAA	ATAATATGGC	TGTTTATTTT	GTAACGGCA	AATTAGGCTC	TGGAAAGACG	3180
CTCGTTAGCG	TTGGTAAGAT	TCAGGATAAA	ATTGTAGCTG	GGTGCAGGAAAT	AGCAACTAAT	3240
CTTGATTAA	GGCTTCAAAAA	CCTCCCGCAA	GTGGGGAGGT	TCGCTAAAC	GCCTCGCGTT	3300
CTTACAATAC	CGGATAAGCC	TTCTATATCT	GATTTGCTTG	CTATTGGCC	CGGTAATGAT	3360
TCCTACGATG	AAAATAAAAAA	CGGCTTGCTT	GTTCTCGATG	AGTGCCTGAC	TTGGTTTAAT	3420
ACCCGTTCTT	GGAATGATAA	GGAAAGACAG	CCGATTATTG	ATTGGTTCT	ACATGCTCGT	3480

AAATTAGGAT GGGATATTAT TTTTCTTGT CAGGACTTAT CTATTGTTGA TAAACAGGGCG	3600
CGTTCTGCAT TAGCTGAACA TGTGTTTAT TGTGTCGTC TGGACAGAAT TACTTTACCT	3660
TTTGTGCGTA CTTTATATTTC TCTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT	3720
GTTGGCGTTG TAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGCG TTGGCTTTAT	3780
ACTGGTAAGA ATTTGTATAA CGCATATGAT ACTAAACAGG CTTTTCTAG TAATTATGAT	3840
TCCGGTGTGTT ATTCTTATTTC AACGCCTTAT TTATCACACG GTCGGTATTT CAAACCATT	3900
AATTTAGGTC AGAAGATGAA GCTTACTAAA ATATATTGA AAAACTTTTC ACGGGTTCTT	3960
TGTCTTGCAGA TTGGATTGTC ATCAGCATT ACATATAGTT ATATAACCCA ACCTAACCG	4020
GAGGTTAAAA AGGTAGTCTC TCAGACCTAT GATTTGATA AATTCACTAT TGACTCTTCT	4080
CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAAT	4140
AGCGACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTTATG TACTGTTCC	4200
ATTAAAAAAAG GTAATTCAAA TGAAATTGTT AAATGTAATT AATTTGTTT TCTTGATGTT	4260
TGTTTCATCA TCTCTTTTG CTCAGGTAAT TGAAATGAAT AATTCGCCTC TGGCCGATTT	4320
TGTAACCTGG TATTCAAAGC AATCAGGCCA ATCCGTTATT GTTTCTCCCG ATGTAAGG	4380
TACTGTTACT GTATATTCACT CTGACGTTAA ACCTGAAAAT CTACGGAATT TCTTTATTTC	4440
TGTTTACGT GCTAATAATT TTGATATGGT TGTTCAATT CCTCCATAA TTCAGAAGTA	4500
TAATCCAAAC AATCAGGATT ATATTGATGA TTGCCATCA TCTGATAATC AGGAATATGA	4560
TGATAATTCC GCTCCTCTG GTGGTTCTT TGTTCCGAA AATGATAATG TTACTCAAAC	4620
TTTTAAAATT AATAACGTTG GGGCAAAGGA TTTAATACGA GTTGTGAAAT TGTTGTAAA	4680
GTCTAAACT TCTAAATCCT CAAATGTATT ATCTATTGAC GGCTCTAAC TATTAGTTGT	4740
TAGTGCACCT AAAGATATT TAGATAACCT TCCTCAATTG CTTTCTACTG TTGATTTGCC	4800
AACTGACCAAG ATATTGATTG AGGGTTGAT ATTTGAGGTT CAGCAAGGTG ATGCTTTAGA	4860
TTTTTCATTT GCTGCTGGCT CTCAGCGTGG CACTGTTGCA GGCGGTGTTA ATACTGACCG	4920
CCTCACCTCT GTTTATCTT CTGCTGGTGG TTGCTCGGT ATTTTAATG GCGATGTTTT	4980
AGGGCTATCA CTTCGCGCAT TAAAGACTAA TAGCCATTCA AAAATATTGT CTGTGCCACG	5040
TATTCTTACG CTTTCAGGTC AGAAGGGTTC TATCTCTGTT GGCCAGAATG TCCCTTTAT	5100
TACTGGTCGT GTGACTGGTG AATCTGCAA TGAAATAAT CCATTGAGA CGATTGAGCG	5160
TCAAAATGTA GGTATTTCCA TGAGCGTTT TCGCTTGGCA ATGGCTGGCG CTAATATTGT	5220
TCTGGATATT ACCACCAAGG CCGATAGTTT GAGTTCTTCT ACTCAGGCAA GTGATGTTAT	5280
TACTAATCAA AGAAGTATTG CTACAACGGT TAATTGCGT GATGGACAGA CTCTTTACT	5340
CGGTGGCCTC ACTGATTATA AAAACACTTC TCAACATTCT GGCGTACCGT TCCTGTCTAA	5400
AATCCCTTAA ATCGGCCTCC TGTTTAGCTC CCGCTCTGAT TCCAACGAGG AAAGCACGTT	5460
ATACGTGCTC GTCAAAGCAA CCATAGTACG CGCCCTGTAG CGGCGCATTAA AGCCGGCGG	5520
GTGTGGTGGT TACCGCGAGC GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCCGCTCCTT	5580

TCGCTTTCTT CCCTTCCTTT CTCGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC	5640
GGGGGCTCCC TTTAGGGTTC CGATTTAGTG CTTTACGGCA CCTCGACCCC AAAAAACTTG	5700
ATTGGGTGA TGGTTACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTGA	5760
CGTTGGACTC CACGTTCTT AATAGTGGAC TCTTGTCCA AACTGGAACA ACACCTCAACC	5820
CTATCTCGGG CTATTCTTT GATTTATAAG GGATTTGCC GATTTCGGAA CCACCATCAA	5880
ACAGGATTTC CGCCTGCTGG GGCAAACCGAG CGTGGACCGC TTGCTGCAAC TCTCTCAGGG	5940
CCAGGGGTG AAGGGCAATC AGCTGTTGCC CGTCTCGCTG GTGAAAAGAA AAACCACCC	6000
GGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGCC GATTCAATTAA TGCAGCTGGC	6060
ACGACAGGTT TCCCAGCTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAAT GTGAGTTAGC	6120
TCACTCAATTA GGCACCCAG GCTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGAA	6180
TTGTGAGCGG ATAACAATTTC CACACGCGTC ACTTGGCACT GGCGTCGTT TTACAACGTC	6240
GTGACTGGGA AAACCCCTGGC GTTACCAAG CTTTGTACAT CGAGAAAATA AAGTGAACAA	6300
AAGCACTATT GCACTGGCAC TCTTACCGTT ACTGTTTACC CCTGTGCCAA AAGCCTATGG	6360
GGGGTTTATG ACTTCTGAGG GATCCGGAGC TGAAGGGAT GACCGCTGCTA AGGCTGCATT	6420
CAATAGTTTA CAGGCAAGTG CTACTGAGTA CATTGGCTAC GCTTGGCTA TGGTAGTAGT	6480
TATAGTTGGT GCTACCATAG CGATTAAATT ATTAAAAAG TTACGAGCA AGGCTTCTTA	6540
ACCAATAGGG AAGAGGCCCG CACCGATCCC CCTTCCCAAC AGTTGCGCAG CCTGAATGGC	6600
GAATGGCCCT TTGGCTGGTT TCCGGCACCA GAAGCGGTGC CGGAAAGCTG GCTGGAGTGC	6660
GATCTTCCCTG AGGCCGATAC GGTGTCGTC CCCTCAAAC GGCAGATGCA CGGTTACGAT	6720
GCGCCCATCT ACACCAACGT AACCTATCCC ATTACGGTCA ATCCGCCGTT TGTCCCACG	6780
GAGAATCCGA CGGGTTGTTA CTCGCTCACA TTAAATGTTG ATGAAAGCTG GCTACAGGAA	6840
GGCCAGACGC GAATTATTT TGATGGCGTT CCTATTGGTT AAAAAATGAG CTGATTTAAC	6900
AAAAATTAA CGGAATTAA AACAAAATAT TAACGTTTAC AATTAAATA TTGCTTATA	6960
CAATCTTCCT GTTTTGGGG CTTTCTGAT TATCAACCGG GGTACATATG ATTGACATGC	7020
TAGTTTACG ATTACCGTTC ATGGATTCTC TTGTTGCTC CAGACTCTCA GGCAATGACC	7080
TGATAGCCTT TGTAGATCTC TCAAAATAG CTACCCCTCTC CGGCATTAAT TTATCAGCTA	7140
GAACGGTTGA ATATCATATT GATGGTGATT TGACTGTCTC CGGGCTTTCT CACCCCTTTG	7200
AATCTTTACC TACACATTAC TCAGGCATTG CATTAAAT ATATGAGGGT TCTAAAAATT	7260
TTTATCCTTG CGTTGAAATA AAGGCTTCTC CGGCAAAAGT ATTACAGGGT CATAATGTTT	7320
TTGGTACAAC CGATTTAGCT TTATGCTCTG AGGCTTTATT GCTTAATTTC GCTAATTCTT	7380
TGCCTTGCCT CTATGATTAA TTGGACGTT	7409

(2) INFORMATION FOR SEQ ID NO:5:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7294 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTICAG CTGGGGCCCC	AAATGAAAAT 60
ATAGCTAAAC AGGTTATTGA CCATTTGCGA AATGTATCTA ATGGTCAAAC	TAAATCTACT 120
CGTTCGCAGA ATTGGGAATC AACTGTTACA TGGAAATGAAA CTTCCAGACA	CCGTAACCTTA 180
CTTGCATATT TAAAACATGT TGAGCTACAG CACCAAGATTAC AGCAATTAAAG	CTCTAAGGCCA 240
TCTGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAGG TACTCTCTAA	TCCTGACCTG 300
TTGGAGTTG CTTCCGGTCT GGTTCGCTT GAAGCTCGAA TTAAAACCGG	ATATTGAAAG 360
TCTTTGGGC TTCCCTTTAA TCTTTTGAT GCAATCCGCT TTGCTTCTGA	CTATAATAGT 420
CACGGTAAAG ACCTGATTT TGATTTATGG TCATTCTCGT TTTCTGAAC	GTAAAGCA 480
TTTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGAG TATTGGACGC	TATCCAGTCT 540
AAACATTTA CTATTACCCC CTCTGGCAAAC ACTTCTTTG CAAAAGCCTC	TCGCTATTTT 600
GGTTTTATC GTCGTCTGGT AAACCGAGGGT TATGATAGTG TTGCTCTTAC	TATGCCCTGGT 660
AATTCCCTTT GGCGTTATGT ATCTGCATTA GTTGAATGTG GTATTCCCTAA	ATCTCAACTG 720
ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA	CGTAGATTTT 780
TCTTCCCAAC GTCCCTGACTG GTATAATGAG CCAGTTCTTA AAATCGCATA	AGGTAAATTCA 840
CAATGATTAAG TTGGAAATT AAACCATCTC AAGCCCAATT TACTACTCCT	TCTGGTGT 900
CTCGTCAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTG TTACGTTGAT	TTGGGTAAATG 960
AATATCCGGT TCTTGTCAAG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT	GCGCCTGGTC 1020
TGTACACCGT TCATCTGTCC TCTTCAAAG TTGGTCAGTT CGGTTCCCTT	ATGATTGACC 1080
GTCTGCGCCT CGTTCCGGCT AAGTAACATG GAGCAGGTGCG CGGATTTCGA	CACAATTTAT 1140
CAGGCGATGA TACAAATCTC CGTTGTACTT TGTTTCCGCG TTGGTATAAT	CGCTGGGGGT 1200
CAAACATGAG TGTTTAGTG TATTCTTTG CCTCTTTCTGT TTAGGTTGG	TGCCTTCGTA 1260
GTGGCATTAC GTATTTTACCGT TTAACTCCCT AAACCTCCTC ATGAAAAAGT	CTTGTACCT 1320
CAAAGCCTCT GTAGCCGTG CTACCCCTCGT TCCGATGCTG TCTTCCGCTG	CTGAGGGTGA 1380
CGATCCCGCA AAAGCGGCCT TTAACTCCCT GCAAGCCTCA GCGACCGAAT	ATATCGGTTA 1440
TGCCGTGGGGCG ATGGTTGTG TCATTGTGG CGCAACTATC GGTATCAAGC	TGTTAAGAA 1500
ATTCACCTCG AAAGCAAGCT GATAAAACCGA TACAATTAAA GGCTCCCTTT	GGAGCCTTTT 1560
TTTTGGAGA TTTTCAACGT GAAAAAATTA TTATTGGCAA TTCCCTTAGT	TGTTCCCTTC 1620
TATTCTCACT CCGCTGAAAC TGTTGAAAGT TGTTAGCAA AACCCCATAC	AGAAAATTCA 1680

TTTACTAACG TCTGGAAAGA CGACAA-LACT TTAGATCGTT ACGCTAACTA TGAGGGTTGT	1740
CTGTGGAATG CTACAGGGCGT TGTAGTTTGT ACTGGTGACG AAACTCAGTG TTACGGTACA	1800
TGGGTTCTA TTGGGCTTGC TATCCCTGAA AATGAGGGTG GTGGCTCTGA GGGTGGCGGT	1860
TCTGAGGGTG GCGGTTCTGA GGGTGGCGGT ACTAAACCTC CTGAGTACGG TGATACACCT	1920
ATTCCGGCT ATACTTATAT CAACCCCTCTC GACGGCACTT ATCCGGCTGG TACTGAGCAA	1980
AACCCCGCTA ATCCTAATCC TTCTCTTGAG GAGTCTCAGC CTCTTAATAC TTTCATGT	2040
CAGAATAATA GGTTCGGAAA TAGGCAGGGG GCATTAACGT TTTATACGGG CACTGTTACT	2100
CAAGGCAGTG ACCCCGTTAA AACTTATTAC CAGTACACTC CTGTATCATC AAAAGCCATG	2160
TATGACGCTT ACTGGAACGG TAAATTAGA GACTGCGCTT TCCATTCTGG CTTTAATGAA	2220
GATCCATTCTG TTTGTGAATA TCAAGGCCAA TCGTCIGACC TGCCTCAACC TCCGTCAAT	2280
GCTGGCGGCG GCTCTGGTGG TGGTTCTGGT GGGGGCTCTG AGGGTGGTGG CTCTGAGGGT	2340
GGCGGTTCTG AGGGTGGCGG CTCTGAGGGA GGGGGTCTCG GTGGTGGCTC TGGTTCCGGT	2400
GATTTGATT ATGAAAAGAT GGCAAAACGCT AATAAGGGGG CTATGACCGA AAATGCCGAT	2460
GAAAACGCGC TACAGTCTGA CGCTAAAGGC AAACTTGATT CTGTCGGTAC TGATTACGGT	2520
GCTGCTATCG ATGGTTTCAT TGGTGACGTT TCCGGCCTTG CTAATGGTAA TGGTGCTACT	2580
GGTGATTTTG CTGGCTCTAA TTCCCAAATG GCTCAAGTCTG CTGACGGTGA TAATTGACCT	2640
TTAATGAATA ATTTCCGTCA ATATTTACCT TCCCTCCGTC AATCGGTGA ATGTCGCCCT	2700
TTTGTCTTAA GCGCTGGTAA ACCATATGAA TTTTCTATTG ATTGTGACAA AATAAACTTA	2760
TTCCGTGGTG TCTTTATAT GTGCCACCT TTATGTATGT ATTTTCTACG	2820
TTTGCTAACAA TACTGCGTAA TAAGGAGTCT TAATCATGCC AGTTCTTTG GGTATTCCGT	2880
TATTATTGCG TTCCCTCGGT TTCCCTCTGG TAACTTTGTT CGGCTATCTG CTTACTTTTC	2940
TTAAAAAGGG CTTCGGTAAG ATAGCTATTG CTATTICATT GTTTCTGCT CTTATTATTG	3000
GGCTTAACTC AATTCTGTG GGTTATCTCT CTGATATTAG CGCTCAATTAA CCCTCTGACT	3060
TTGTTCAAGGG TGTTCAGTTA ATTCTCCCGT CTAATGCGCT TCCCTGTTT TATGTTATTG	3120
TCTCTGTAAA GGCTGCTATT TTCACTTTTG ACGTTAAACA AAAAATCGTT TCTTATTGTTG	3180
ATTGGGATAA ATAATATGGC TGTTTATTGTT GTAACTGGCA AATTAGGCTC TGGAAAGAGG	3240
CTCGTTAGCG TTGGTAAGAT TCAGGATAAA ATTCTAGCTG GGTGCAAAAT AGCAACTAAT	3300
CTTGATTAA GGCTTCAAAA CCTCCCGCAA GTCGGGAGGT TCGCTAAAC GCCTCGCGTT	3360
CTTAGAATAC CGGATAAGCC TTCTATATCT GATTGCTTG CTATTGGCG CGCTAAATGAT	3420
TCCTACGATG AAAATAAAA CGGCTTGCTT GTTCTCGATG AGTGCCTAC TTGGTTTAAT	3480
ACCCGTTCTT GGAATGATAA GGAAAGACAG CCGATTATTG ATTGGTTCT ACATGCTCGT	3540
AAATTAGGAT CGGATATTAT CTTCCTGTT CAGGACTTAT CTATTGTTGA TAAACAGGCG	3600
CGTTCTGCAT TAGCTGAACA TGTGTTTAT TGTCTCGTC TGGACAGAAAT TACTTTACCT	3660
TTTGTGGTA CTTTATATTG TCTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT	3720

GTTGGCGTTG TTAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGCG TTGGCTTAT	3780
ACTGGTAAGA ATTTGTATAA CGCATATGAT ACTAACACAGG CTTTTCTAG TAATTATGAT	3840
TCCGGTGTTC ATTCTTATTAA AACGCCCTAT TTATCACACAGG GTCCGTATTT CAAACCATTA	3900
AATTAGGTC AGAAGATGAA GCTTACTAAA ATATATTGA AAAAGTTTC ACGCGTTCTT	3960
TGTCTTGCAG TTGGATTTCG ATCAGCATT ACATATAGTT ATATAACCCA ACCTAAGCCG	4020
GAGGTTAAAA AGGTAGTC TCAGACCTAT GATTTGATA AATTACTAT TGACTCTTCT	4080
CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAAAT	4140
AGCCACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTATG TACTGTTCC	4200
ATTAAAAAGG TAATTCAAAT GAAATTGTTA AATGTAATT ATTTTGTTTT CTTGATGTTT	4260
GTTTCATCAT CTTCTTTGC TCAGGTAATT GAAATGAATA ATTCCGCTCT GCGCGATTTT	4320
GTAACTTGGT ATTCAAAGCA ATCAGGCAGA TCCGTTATTG TTTCTCCGA TCTAAAAGGT	4380
ACTGTTACTG TATATTCATC TGACGTTAAA CCTGAAAATC TACGCAATTTC CTTTATTCT	4440
GTTTACGTG CTAATAATTAA TGATATGGTT GGTTCAATTTC CTTCCATTAT TTAGAACTAT	4500
AATCCAAACA ATCAGGATTAA TATTGATGAA TTGCCATCAT CTGATAATCA GGAATATGAT	4560
GATAATTCCG CTCCCTCTGG TGGTTCTTT GTTCCGCAAA ATGATAATGT TACTCAAAC	4620
TTTAAAATTAA ATAACGTTCG GCAAAGGAT TIAATACGAG TTGTCGAATT GTTGTAAAG	4680
TCTAATACCTT CTAATCCTC AAATGTATT TCTATTGACC GCTCTAATCT ATTAGTTGTT	4740
AGTGCACCTA AAGATATTAA AGATAACCTT CCTCAATTCC TTTCTACTGT TGATTGCCA	4800
ACTGACCAGA TATTGATTGA GGGTTGATA TTTGAGGTT AGCAAGGTGA TGCTTCTAGAT	4860
TTTTCATTTG CTGCTGGCTC TCAGCGTGGC ACTGTTGCAG GCGGTGTAA TACTGACCGC	4920
CTCACCTCTG TTTTATCTTC TGCTGGTGGT TCGTTGGTA TTTTAATGG CGATGTTTTA	4980
GGGCTATCAG TTGGCGCATT AAAGACTAAT AGCCATTCAA AAATATTGTC TGTGCCACGT	5040
ATTCTTACGC TTTCAGGTCA GAAGGGTCT ATCTCTGTG GCCAGAATGT CCCTTTTATT	5100
ACTGGTGTG TGACTGGTGA ATCTGCCAAT GTAAATAATC CATTGAGAC GATTGAGCGT	5160
CAAAATGTAG GTATTCCAT GAGCGTTTT CCTGTTGCAA TGGCTGGCGG TAATATTGTT	5220
CTGGATATTA CCAGGAAGGC CGATAGTTG AGTTCTCTA CTCAGCCAAG TGATGTTATT	5280
ACTAATCAA GAAGTATTGC TACAACGGTT AATTGCGTC ATGGACAGAC TCTTTACTC	5340
GGTGGCCTCA CTGATTATAA AAACACTTCT CAAGATTCTG CGTACCGTT CCTGTCTAAA	5400
ATCCCTTAA TCGGCCCTCC GTTTAGCTCC CGCTCTGATT CCAAGGAGGA AAGCACGTTA	5460
TACGTGCTCG TCAAAGCAAC CATACTACGC GCCCTGTAGC GGCGCATTAA GCGCGGGGG	5520
TGTGGTGGTT ACGCGCAGCG TGACCGCTAC ACTTGGCAGC GCCCTAGCGC CCGCTCCTT	5580
CGCTTTCTTC CCTTCCCTTC TCGCCACGTT CGCCGGCTTT CCCCCGTCAAG CTCTAAATCG	5640
GGGGCTCCCT TTAGGGTTCC GATTTAGTGC TTTACGGCAC CTCGACCCCCA AAAAACTTGA	5700
TTTGGGTGAT CGTTCACGTA GTGGGCCATC GCCCTGATAG ACGGTTTTTC GCCCTTGAC	5760

GTGGAGTCC ACGTTCTTAA ATAGTGGACT CTGTTCCAA ACTGGAACAA CACTCAACCC	5820
TATCTCGGGC TATTCTTTG ATTTATAAGG GATTTGCCG ATTCGGAAC CACCATCAA	5880
CAGGATTTTC GCCTGCTGGG GCAAACCCAGC GTGGACCGCT TGCTGCAACT CTCTCAGGGC	5940
CAGGGGTGA AGGGCAATCA GCTGTTGCCG TGCTCGCTGG TGAAAAGAAA AACCCACCC	6000
GCGCCCAATA CGCAAACCCGC CTCTCCCCGC CGCTTGGCCG ATTCAATTAA GCAGCTGGCA	6060
CGACAGGTTT CCCGACTGGA AAGCCGCCAG TGAGCGAAC GCAATTAAATG TGAGTTAGCT	6120
CACTCATTAG GCACCCCGAGG CTTTACACTT TATGCTTCCG GCTCGTATGT TGTGTGGAAT	6180
TGTGAGCGGA TAACAATTTC ACACAGGAAA CAGCTATGAC CAGGATGTAC GAATTCCAG	6240
GTAGGAGAGC TCGCGGATC CGAGGCTGAA GGCATGACC CTGCTAAGGC TGCATTCAAT	6300
AGTTTACAGG CAAAGTGTAC TGAGTACATT GGCTACGCTT GGGCTATGGT AGTAGTTATA	6360
GTTGGTGCTA CCATAGGGAT TAAATTATTC AAAAAGTTA CGAGCAAGGC TTCTTAACCA	6420
GCTGGCGTAA TAGCGAAGAG GCCCCCACCG ATCGCCCTTC CCAACAGTTG CGCAGGCTGA	6480
ATGGCGAATG GCGCTTGCC TCGTTTCCGG CACCAGAAGC GGTGCCGAA AGCTGGCTGG	6540
AGTGGCATCT TCCTGAGGCC GATACTGGTCG TCGTCCCTC AAACTGGCAG ATGCACGGTT	6600
ACGATGCCGC CATCTACACC AACGTAACCT ATCCCATTAAC GGTCAATCCG CCGTTTGTTC	6660
CCACGGAGAA TCCGACGGGT TGTTACTCGC TCACATTAA TGTGATGAA AGCTGGCTAC	6720
AGGAAGGCCA GACCGAATT ATTTTGATC CGTTCCCTAT TGGTTAAAAA ATGAGCTGAT	6780
TTAACAAAAA TTAAACCGGA ATTTAACAA AATATTAACG TTTACAATTAA AAATATTGCG	6840
TTATACAATC TTCTGTTTT TGGGGCTTT CTGATTATCA ACCGGGGTAC ATATGATTGA	6900
CATGCTAGTT TTACGATTAC CGTCATCGA TTCTCTTGTGTT TGCTCCAGAC TCTCAGGCAA	6960
TGACCTGATA GCCTTGTAG ATCTCTAAA AATAGCTACC CTCTCCGGCA TTAATTATC	7020
AGCTAGAACG GTTGAATATC ATATTGATGG TGATTGACT GTCTCCGGCC TTTCTCACCC	7080
TTTTGAATCT TTACCTACAC ATTACTCAGG CATTGGATT AAAATATATG AGGGTTCTAA	7140
AAATTCTTAT CCTTGGTTG AAATAAGGC TTCTCCGGCA AAAGTATTAC AGGGTCATAA	7200
TGTTTTGGT ACAACCGATT TAGCTTATG CTCTGAGGCT TTATTGCTTA ATTTTGCTAA	7260
TTCTTGCCT TGCCTGTATG ATTTATTGGA CGTT	7294

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7394 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTCAAG CTCGGCCCC AAATGAAAAT

ATAGCTAAC AGGTTATTGA CCATTTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT	120
CGTTCCGAGA ATTGGGAATC AACTGTTACA TGGAATGAAA CTTCCAGACA CCGTACTTTA	180
GTTGCATATT TAAAACATGT TGAGCTACAG CACCAGATTG AGCAATTAAG CTCTAAGCCA	240
TCTGCAAAAA TGACCTCTTA TCAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG	300
TTGGAGTTG CTTCCGGTCT GGTCGCTT GAAGCTCGAA TTAAAACGCG ATATTTGAAG	360
TCTTCGGGC TTCCCTCTAA TCTTTTGAT GCAATCCGCT TTGCTCTGA CTATAATAGT	420
CAGGGTAAAG ACCTGATTTC TGATTTATGG TCATTCTCGT TTTCTGAACG GTTAAAGCA	480
TTTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT	540
AAACATTTA CTATTACCCC CTCTGGAAA ACTTCTTTG CAAAAGCCTC TCGCTATTT	600
GGTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC TATGCCCTCGT	660
AATTCCCTTT GGCGTTATGT ATCTGCATTA GTTGAATGTG GTATTCCTAA ATCTCAACTG	720
ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA CGTAGATTT	780
TCTTCCCAAC GTCCCTGACTG GTATAATGAG CCAGTTCTTA AAATCGCATA AGGTAATTCA	840
CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGT	900
CTCGTCAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTG TTACGTTGAT TTGGGTAATG	960
AATATCCGGT TCTTGTCAAG ATTACTCTG ATGAAGGTCA GCCAGCCTAT GCGCCTGGTC	1020
TGTACACCGT TCATCTGTCC TCTTCAAAG TTGGTCAGTT CGGTTCCCTT ATGATTGACC	1080
GTCTGCCCT CGTTCCGGCT AAGTAACATG GAGCAGGTGG CGGATTTCGA CACAATTAT	1140
CAGCGATGA TACAAATCTC CGTTGTACTT TGTTTGGCC TTGGTATAAT CGCTGGGGGT	1200
CAAAGATGAG TGTTTTAGTG TATTGTTCG CCTCTTCTGT TTTAGGTTGG TGCCTTCGTA	1260
GTGGCATTAC GTATTTTAC CGTTAATGG AAACCTCCTC ATGAAAAAGT CTTAGTCCT	1320
CAAAGCCTCT GTAGCCGTTG CTACCCCTCGT TCCGATGCTG TCTTCGCTG CTGACGGTGA	1380
CGATCCCGCA AAAGCGGCCT TTAACCTCCCT GCAAGCCTCA CGCACCGAAT ATATCGTTA	1440
TCCGTGGCG ATGGTTGTTG TCATTGTCGG CGCAACTATC GGTATCAAGC TGTTAAGAA	1500
ATTCACCTCG AAAGCAAGCT GATAAACCGA TACAATTAAA GGCTCCTTT GGAGCCTTT	1560
TTTTGGAGA TTTTCAACGT GAAAAAAATTA TTATTCGAA TTGCTTTAGT TGTTCTTTC	1620
TATTCTCACT CCGCTGAAAC TGTTGAAAGT TGTTAGCAA AACCCCATAC AGAAAATTCA	1680
TTTACTAACG TCTGGAAAGA CGACAAAAGT TTAGATCGTT ACGCTAACTA TGAGGGTTGT	1740
CTGTGGAATG CTACAGGGCT TCTAGTTGT ACTGGTGACG AAAACTCAGTG TTACGGTACA	1800
TGGGTTCCCTA TTGGGCTTGC TATCCCTGAA AATGAGGGTG GTGGCTCTGA GGGTGGCGGT	1860
TCTGAGGGTG GCGGTTCTGA GGGTGGCGGT ACTAAACCTC CTGAGTACGG TGATACACCT	1920
ATTCCGGGCT ATACTTATAT CAACCCCTCTC GACGGCACTT ATCCGCCCTGG TACTGAGCAA	1980
AACCCCGCTA ATCCTAAATCC TTCTCTTGAG GAGTCTCAGC CTCTTAATAC TTTCATGTTT	2040
CAGAATAATAA GGTTCCGAAA TAGGCAGGGG GCATTAACGTG TTTATACGGG CACTGTTACT	2100

CAAGGCAC TG ACCCGTTAA AACTTATTAC CAGTACACTC CTGTATC ATC AAAAGCCATG	2160
TATGACCGCTT ACTGGAACGG TAAATT CAGA GACTGCCCTT TCCATTCTGG CTTTAATGAA	2220
GATCCATT CG TTTGTGAATA TCAAGGCCAA TCGTCTGACC TGCCCTCAACC TCCGTGCAAT	2280
GCTGGCGGCG GCTCTGGTGG TGGTTCTGGT GCCGGCTCTG AGGGTGGTGG CTCTGAGGGT	2340
GCGCGTCTG AGGGTGGCGG CTCTGAGGGA GCCGGTTCCCG GTGCGTGGCTC TGGTTCCGGT	2400
GATTTGATT ATGAAAAGAT GCGAAACGCT AATAAGGGGG CTATGACCGA AAATGCCGAT	2460
GAAAACGCGC TACAGTCTGA CGCTAAAGGC AAACCTGATT CTGTCGCTAC TGATTACGGT	2520
GCTGCTATCG ATGGTTTCAT TGGTGACGTT TCCGGCCTTG CTAATGGTAA TGGTGCTACT	2580
GGTGATTTTG CTGGCTCTAA TTCCCAAATG GCTCAAGTCC GTGACGGTGA TAATTCA CCT	2640
TTAATGAATA ATTTCCGTCA ATATTACCT TCCCTCCCTC AATCGGTGA ATGTCGCCCT	2700
TTTGTCTTAA GCGCTGGTAA ACCATATGAA TTTTCTATTG ATTGTGACAA AATAAACTTA	2760
TTCCGTGGTG TCTTTGCGTT TCTTTATAT GTGCCACCT TTATGTATGT ATTTTCTACG	2820
TTTGCTAAC A TACTGCGTAA TAAGGAGTCT TAATCATGCC AGTTCTTTG GGTATTCCGT	2880
TATTATTGCG TTTCCTCGGT TTGCTTCTGG TAACCTTGTG CGGCTATCTG CTTACTTTTC	2940
TTAAAAAGGG CTTCGGTAAG ATAGCTATTG CTATTCATT GTTCTTGCT CTTATTATTG	3000
GGCTTAAC TC AATTCTTG GCTTATCTCT CTGATATTAG CGCTCAATTAA CCCTCTGACT	3060
TTGTTCAAGG TGTCAGTTA ATTCTCCGT CTAATGCCGT TCCCTGTTT TATGTTATTC	3120
TCTCTGTAAA GGCTGCTATT TTCACTTTTG ACGTTAAACA AAAAATCGTT TCTTATTG	3180
ATTGGGATAA ATAATATGGC TGTTTATTT GTAACTGGCA AATTAGGCTC TGGAAAGACG	3240
CTCGTTAGCG TTGCTAAGAT TTAGGATAAA ATTGTAGCTG GGTGCAAAT AGCAACTAAT	3300
CTTGATTAA GGCTCAAAA CCTCCGGAA GTCGGGAGGT TCGCTAAAC GCCTCGCGTT	3360
CTTAGAATAC CGGATAAGCC TTCTATATCT GATTGCTTG CTATTGGCG CGGTAAATGAT	3420
TCCTACGATG AAAATAAAA CGGCTTGCCTT GTTCTCGATG AGTGGGTAC TTGGTTAAAT	3480
ACCCGTTCTT GGAATGATAA GGAAAGACAG CCGATTATTG ATTGGTTCT ACATGCTCGT	3540
AAATTAGGAT GGGATATTAT TTCTCTGTT CAGGACTTAT CTATTGTTGA TAAACAGGGC	3600
CGTTCTGCAT TAGCTGAACA TGTTGTTAT TGTCGTCGTC TGGACAGAAT TACTTTACCT	3660
TTTGTGGTA CTTTATATTC TCTTATTACT GGCTCGAAAA TGCCCTGCC TAAATTACAT	3720
GTGGCGTTG TTAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGCG TTGGTTTAT	3780
ACTGGTAAGA ATTGTATAA CGCATATGAT ACTAAACAGG CTTTTCTAG TAATTATGAT	3840
TCGGGTGTTT ATTCTTATT AAGCCTTAT TTATCACACG GTCCGTATTT CAAACCATTA	3900
AATTAGGTC AGAAGATGAA GCTTACTAAA ATATATTGA AAAAGTTTC ACGCGTTCTT	3960
TGTCTTGGCA TTGGATTTCG ATCAGCATT ACATATAGTT ATATAACCCA ACCTAACCGG	4020
GAGGTAAAA AGGTAGTCTC TCAGACCTAT GATTTGATA AATTCACTAT TGACTCTTCT	4080
CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAAT	4140

AGCGACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTATG TACTGTTCC	4200
ATTAAAAAAAG GTAATTCAAA TGAAATTGTT AAATGTAATT AATTTGTTT TCTTGATGTT	4260
TGTTTCACTCA TCCTCTTTG CTCAGGTAAT TGAAATGAAT AATTGGCTC TGGCGGATT	4320
TGTAACCTGG TATTCAAAGC AATCAGGCGA ATCCGTTATT GTTCTCCCG ATGTAAAAGG	4380
TACTGTTACT GTATATTCACT TGACGTTAA ACCTGAAAT CTACGCAATT TCTTTATTT	4440
TGTTTACGT GCTAATAATT TTGATATGGT TGGTTCAATT CCTTCCATAA TTCAGAAGTA	4500
TAATCCAAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATGA	4560
TGATAATTCC GCTCCTCTG GTGGTTCTT TGTCCGCAA AATGATAATG TTACTCAAAC	4620
TTTAAAAATT AATAACGTTC GGGCAAAGGA TTTAATACGA GTTGTGAAT TGTGTAAA	4680
GTCTAATACT TCTAAATCCT CAAATGTATT ATCTATTGAC GGCTCTAAC TATTAGTTGT	4740
TAGTGCACCT AAAGATATT TAGATAACCT TCCTCAATC CTTTCTACTG TTGATTTGCC	4800
AACTGACCAAG ATATTGATTG AGGGTTGAT ATTTGAGGTT CAGCAAGGTG ATGCTTTAGA	4860
TTTTGATTT GCTGCTGGCT CTCAGCGTGG CACTGTTCCA GGCGGTGTTA ATACTGACCG	4920
CCTCACCTCT GTTTTATCTT CTGCTGGTGG TTGCTTCGGT ATTTTTAATG GCGATGTTTT	4980
AGGGCTATCA GTTCCGCAT TAAAGACTAA TAGCCATTCA AAAATATTGT CTGTGCCACG	5040
TATTCTTAGG CTTTCAGGTC AGAAGGGTTC TATCTCTGTT GGCCAGAATG TCCCTTTAT	5100
TACTGGTCGT GTGACTGGTG AATCTGCCAA TGAAATAAT CCATTTAGA CGATTGAGCG	5160
TCAAAATGTA GGTATTTCGA TGAGCGTTT TCCTGTTGCA ATGGCTGGGG GTAATATTGT	5220
TCTGGATATT ACCAGCAAGG CCGATAGTTT GAGTTCTCT ACTCAGGCAA GTGATGTTAT	5280
TACTAATCAA AGAAGTATTG CTACAACGGT TAATTGCGT GATGGACAGA CTCTTTACT	5340
CGGTGGCCTC ACTGATTATA AAAACACTTC TCAAGATTCT GGCGTACCGT TCCTGTCTAA	5400
AATCCCTTTA ATCGGCCTCC TGTTAGCTC CCGCTCTGAT TCCAACGAGG AAAGCACGTT	5460
ATACGTGTC GTCAGGCAA CCATAGTACG CGCCCTGTAG CGGGCGATTAA AGCCGGGGGG	5520
GTGTGGTGGT TACGGCGAGC GTGACGGCTA CACTGCCAG CGCCCTAGCG CCCGCTCCTT	5580
TCGCTTCTT CCCTTCCTT CTCGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC	5640
GGGGGCTCCC TTAGGGTTC CGATTTAGTG CTTCAGGCA CCTCGACCCC AAAAAACTTG	5700
ATTGGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTTGA	5760
CGTTGGAGTC CACGTTCTT AATAGTGGAC TCTGTTCCA AACTGGAACA ACACCTCAACC	5820
CTATCTCGGG CTATTCTTT GATTATAAG GGATTTGCC GATTTGGAA CCACCATCAA	5880
ACAGGATTAA CGCCTGCTGG GGCAAACCAAG CGTGGACCGC TTGCTGCAAC TCTCTCAGGG	5940
CCAGGGGGTG AAGGGCAATC AGCTGTTGCC CGTCTCGCTG GTGAAAAGAA AAACCACCC	6000
GGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGCC GATTCACTAA TGCAGCTGGC	6060
ACGACAGGTT TCCCCACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC	6120
TCACTCACTTA GGCACCCCCAG CCTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA	6180

TTGTGAGCGG ATAACAATT CACACGGC ACCTGGCACT GGCGTGGTT TTACAACGTC	6240
GTGACTCGGA AAACCCCTGGC GTTACCCAAG CTTTGTACAT GGAGAAAATA AACTGAAACA	6300
AAGCACTATT GCACCTGGC ACCTTACCGTT ACTGTTTACG CCTGTGGCAA AAGCCCTTCT	6360
GAGGCATCCG GGAGCTGAAG GCGATGACCC TGCTAAGGCT GCATICAATA GTTACAGGC	6420
AAGTGTACT GAGTACATTG GCTACGGTTG GGCTATGGTA GTAGTTATAG TTGGTGTAC	6480
CATAGGGATT AAATTATTCA AAAAGTTTAC GAGCAAGGCT TCTTAAGCAA TAGCGAAGAG	6540
GCCCGCACCG ATCGCCCTTC CCAACAGTTG CGCAGCCTGA ATGGCGAATG GCGCTTGCC	6600
TGGTTTCCGG CACCAGAACG GGTGGGGAA AGCTGGCTGG AGTGCATCT TCCTGAGGCC	6660
GATACGGTCG TCGTCCCCTC AAACTGGCAG ATGCACGGTT ACGATGGCC CATCTACACC	6720
AACGTAACCT ATCCCATTAC GGTCAATCCG CCGTTTGTTC CCACGGAGAA TCCGACGGGT	6780
TGTTACTCGC TCACATTAA TGTTGATGAA AGCTGGCTAC AGGAAGGCCA GACCGGAATT	6840
ATTTTGATG CGCTTCCTAT TGGTTAAAAA ATGAGCTGAT TTAACAAAAA TTTAACGGGA	6900
ATTTTAACAA AATATTAACG TTTACAATT AAATATTTCG TTATACAATC TTCTGTTTT	6960
TGGGGTTTT CTGATTATCA ACCGGGGTAC ATATGATTGA CATGCTAGTT TTACGATTAC	7020
CGTTCATCGA TTCTCTTGTG TGCTCCAGAC TCTCAGGCAA TGACCTGATA GCCTTGTAG	7080
ATCTCTCAA AATAGCTACC CTCTCCGGCA TTAATTATAC AGCTAGAACG GTTGAATATC	7140
ATATTGATGG TGATTTGACT GTCTCCGGCC TTCTCACCC TTTGAATCT TTACCTACAC	7200
ATTACTCAGG CATTGCATT AAAATATATG AGGGTTCTAA AAATTTTAT CCTTGCCTTG	7260
AAATAAAGGC TTCTCCGGCA AAAGTATTAC AGGGTCATAA TGTTTTGGT ACAACCGATT	7320
TAGCTTTATG CTCTGAGGCT TTATTGCTAA TTCTTGCCT TGCCTGTATG	7380
ATTTATTGGA CGTT	7394

(2) INFORMATION FOR SEQ ID NO:7:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GATCCTAGGC TGAAGGGCAT GACCCTGCTA AGGCTGC

37

(2) INFORMATION FOR SEQ ID NO:8:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

93

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ATTCAATAGT TTACAGGCAA GTGCTACTGA GTACA

35

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TTGGCTACCC TTGGGCTATG GTAGTAGTTA TAGTT

35

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GGTGCTACCA TAGGGATTAA ATTATTCAAA AAGTT

35

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TACGAGCAAG GCTTCTTA

18

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

AGCTTAAGAA GCCTTGCTCG TAAACTTTT GAATAATT

39

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AATCCCTATG GTAGCACCAA CTATAACTAC TACCAT

36

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AGCCAAGCG TAGCCAATGT ACTCAGTAGC ACTTG

35

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CCTGTAAACT ATTGAATGCA GCCTTAGCAG GGTC

34

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

ATGGCCTTCA GCCTAG

16

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CTCGAATTCTG TACATCCTGG TCATAGC

27

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CATTTTGCA GATGGCTTAG A

21

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TAGCATTAAAC GTCCAATA

18

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

ATATATTTTA GTAAGCTTCA TCTTCT

26

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GACAAAGAAC GCGTGAAAAC TTT

23

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GGGGGCCTCT TCGCTATTGC TTAAGAAGCC TTGCT

35

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TTCAGCCTAG GATCCGCCGA GCTCTCCTAC CTGCGAATTC GTACATCC

43

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

TGGATTATAC TTCTAAATAA TGGA

24

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TAACACTGAT TCCGGATGGA ATTCTGGAGT CTGGGT

36

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

AATTGCCAA GGAGACAGTC AT

22

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

AATGAAATAC CTATTGCCTA CGGCAGCCGC TGGATTGTT

39

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

ATTACTCGCT GCCCAACCAAG CCATGGCCGA CCTCGTGAT

39

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GACCCAGACT CCAGATATCC AACAGGAATG AGTGTAAAT

39

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

TCTAGAACGC GTC

13

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

ACGTGACCGCG TTCTAGAATT AACACTCATT CCTGT

35

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TGGATATCTG GAGTCTGGGT CATCACGAGC TCGGCCATG

36

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GCTGGTTGGG CAGCGACTAA TAACAATCCA GCGGCTGCC

39

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GTAGGCAATA CGTATTTCAT TATGACTGTC CTTGGCG

37

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

TGACTGTCTC CTTGGCGTGT GAAATTGTTA

30

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

TAACACTCAT TCCGGATGGA ATTCTGGAGT CTGGGT

36

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CAATTTTATC CTAATCTTA CCAAC

25

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CATTTTTGCA CATGGCTTAG A

21

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CGAAAGGGGG GTGTGCTGCA A

21

100

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

TAGCATTAAAC GTCCAATA

18

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

AAACCGACGGC CAGTGCCAAG TGACGGGTGT GAAATTGTTA TCC

43

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GGCGAAAGGG AATTCTGCAA GGCGATTAAAG CTTGGGTAAAC CCC

43

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GGCGTTACCC AAGCTTGTATGGAGAAA ATAAAG

36

101

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

TGAAACAAAG CACTATTGCA CTGGCACTCT TACCGTTACC CT

43

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

TACTGTTTAC CCCTGTGACA AAAGCCGCC AGGTCCAGCT GC

4.2

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

TCGAGTCAGG CCTATTGTGC CCAGGGATTG TACTAGTGGAA TCCG

4.4

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

TGGCGAAAGG GAATTCCGGAT CCACTAGTAC AATCCCTG

38

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:
GGCACACAATAG GCCTGACTCG AGCAGCTGGA CCAGGGGGC TT

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:
TTGTCACAGG GGTAAACAGT AACGGTAACG GTAAGTGTGC CA

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:
GTGCAATACT GCTTTGTTTC ACTTTATTTT CTCCATGTAC AA

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:
TAACGGTAAG AGTGCCAGTG C

(52) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 68 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(25, "")
(D) OTHER INFORMATION: /note- "M REPRESENTS AN EQUAL
MIXTURE OF A AND C AT THIS LOCATION AND AT
LOCATIONS 28, 31, 34, 37, 40, 43, 46 & 49"

103

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

AGCTCCCGGA TGCCTCAGAA GATGNNNNNN NNNNNNNNNNN NNNNNNNNNNN NGGCTTTGCG
60
CACAGGGG
68

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 54 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(17, "")
- (D) OTHER INFORMATION: /note- "M REPRESENTS AN EQUAL
MIXTURE OF A AND C AT THIS LOCATION AND AT
LOCATIONS 20, 23, 26, 29, 32, 35, 38, 41, 44 & 50"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

CAGCCTCGGA TCCGCCMNNM NNMNNMNNNM NMNNMNNNNNN MNMMNNATGM GAAT

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

GGTAAACAGT AACGGTAAGA GTGCCAG

27

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

GGGCTTTGCG CACAGGGT

19

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 63 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

WO 92/06176

104

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:
 AGGGTCATCG CCTTCAGCTC CGGATCCCTC AGAAGTCATA AACCCCCCAT AGGCTTTGCC 60
 63

CAC

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 47 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:
 TCGCCTTCAG CTCCCGGATG CCTCAGAAGC ATGAACCCCC CATAGGC

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 25 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:
 CAATTTTATC CTAAATCTTA CCAAC 25

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:
 GCCTTCAGCC TCGGATCCGC C 21

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:
 CGGATGCCTC AGAAGCCCCN N 21

105

(2) INFORMATION FOR SEQ ID NO:61:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

CGGATGCCTC AGAAGGGCTT TTGCCACAGG

30

I CLAIM:

1. A composition of matter comprising a plurality of cells containing a diverse population of expressible oligonucleotides operationally linked to expression elements, said expressible oligonucleotides having a desirable bias of random codon sequences produced from random combinations of first and second oligonucleotide precursor populations having a desirable bias of random codon sequences.

2. The composition of claim 1, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is unbiased.

3. The composition of claim 1, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is biased toward a predetermined sequence.

4. The composition of claim 1, wherein said first and second oligonucleotides having random codon sequences have at least one specified codon at a predetermined position.

5. The composition of claim 1, wherein said cells are prokaryotes.

6. The composition of claim 1, wherein said cells are E. coli.

7. A kit for the preparation of vectors useful for the expression of a diverse population of random peptides from combined first and second oligonucleotides having a desirable bias of random codon sequences, 5 comprising: two vectors: a first vector having a cloning site for said first oligonucleotides and a pair of restriction sites for operationally combining first oligonucleotides with second oligonucleotides; and a second vector having a cloning site for said second 10 oligonucleotides and a pair of restriction sites complementary to those on said first vector, one or both vectors containing expression elements capable of being operationally linked to said combined first and second oligonucleotides.

8. The kit of claim 7, wherein said vectors are in a filamentous bacteriophage.

9. The kit of claim 8, wherein said filamentous bacteriophage are M13.

10. The kit of claim 7, wherein said vectors are plasmids.

11. The kit of claim 7, wherein said vectors are phagemids.

12. The kit of claim 7, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is unbiased.

13. The kit of claim 7, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is diverse but biased toward a predetermined sequence.

14. The kit of claim 7, wherein said first and second oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.

15. The kit of claim 7, wherein said pair of restriction sites are Fok I.

16. A cloning system for expressing random peptides from diverse populations of combined first and second oligonucleotides having a desirable bias of random codon sequences, comprising: a set of first vectors having a diverse population of first oligonucleotides having a desirable bias of random codon sequences and a set of second vectors having a diverse population of second oligonucleotides having a desirable bias of random codon sequences, said first and second vectors each 5 having a pair of restriction sites so as to allow the 10 operational combination of first and second oligonucleotides into a contiguous oligonucleotide having a desirable bias of random codon sequences.

17. The cloning system of claim 16, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is unbiased.

18. The cloning system of claim 16, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is diverse but biased toward a predetermined sequence.

19. The cloning system of claim 16, wherein said first and second oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.

20. The cloning system of claim 16, wherein said combined first and second vectors is through a pair of restriction sites.

21. The cloning system of claim 16, wherein said pair of restriction sites are Fok I.

22. A composition of matter comprising a plurality of cells containing a diverse population of expressible oligonucleotides operationally linked to expression elements, said expressible oligonucleotides having a desirable bias of random codon sequences.

23. The composition of claim 22, wherein said cells are prokaryotes.

24. The composition of claim 22, wherein said expressible oligonucleotides are expressed as peptide fusion proteins on the surface of a filamentous bacteriophage.

25. The composition of claim 22, wherein said filamentous bacteriophage is M13.

26. The composition of claim 22, wherein said fusion protein contains the product of gene VIII.

27. The composition of claim 22, wherein said diverse population of oligonucleotides having a desirable bias of random codon sequences are produced from the combination of diverse populations of first and second oligonucleotides having a desirable bias of random codon sequences.

28. The composition of claim 22, wherein the desirable bias of random codon sequences of said oligonucleotides is unbiased.

29. The composition of claim 22, wherein the desirable bias of random codon sequences of said oligonucleotides is diverse but biased toward a predetermined sequence.

30. The composition of claim 22, wherein said oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.

31. A plurality of vectors containing a diverse population of expressible oligonucleotides having a desirable bias of random codon sequences.

32. The vectors of claim 31, wherein said oligonucleotides are expressible as fusion proteins on the surface of filamentous bacteriophage.

33. The vectors of claim 31, wherein said filamentous bacteriophage is M13.

34. The vectors of claim 31, wherein said fusion protein contains the product of gene VIII.

35. The vectors of claim 31, wherein the desirable bias of random codon sequences of said oligonucleotides is unbiased.

36. The vectors of claim 31, wherein the desirable bias of random codon sequences of said oligonucleotides is diverse but biased toward a predetermined sequence.

37. The vectors of claim 31, wherein said oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.

38. A composition of matter, comprising a diverse population of oligonucleotides having a desirable bias of random codon sequences produced from random combinations of two or more oligonucleotide precursor 5 populations having a desirable bias of random codon sequences.

39. A method of constructing a diverse population of vectors having combined first and second oligonucleotides having a desirable bias of random codon sequences capable of expressing said combined 5 oligonucleotides as random peptides, comprising the steps of:

10 (a) operationally linking sequences from a diverse population of first oligonucleotides having a desirable bias of random codon sequences to a first vector;

15 (b) operationally linking sequences from a diverse population of second oligonucleotides having a desirable bias of random codon sequences to a second vector; and

20 (c) combining the vector products of steps (a) and (b) under conditions where said populations of first and second oligonucleotides are joined together into a population of combined vectors capable of being expressed.

40. The method of claim 39, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is unbiased.

41. The method of claim 39, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is diverse but biased toward a predetermined sequence.

42. The method of claim 39, wherein said first and second oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.

43. The method of claim 38, wherein steps (a) through (c) are repeated two or more times.

44. A method of selecting a peptide capable of being bound by a ligand binding protein from a population of random peptides, comprising:

5

(a) operationally linking a diverse population of first oligonucleotides having a desirable bias of random codon sequences to a first vector;

10

(b) operationally linking a diverse population of second oligonucleotides having a desirable bias of random codon sequences to a second vector;

15

(c) combining the vector products of steps (a) and (b) under conditions where said populations of first and second oligonucleotides are joined together into a population of combined vectors;

20

(d) introducing said population of combined vectors into a compatible host under conditions sufficient for expressing said population of random peptides; and

(e) determining the peptide which binds to said ligand binding protein.

45. The method of claim 44, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is unbiased.

46. The method of claim 44, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is diverse but biased toward a predetermined sequence.

47. The method of claim 44, wherein said first and second oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.

48. The method of claim 44, wherein steps (a) through (c) are repeated two or more times.

49. A method for determining the nucleic acid sequence encoding a peptide capable of being bound by a ligand binding protein which is selected from a population of random peptides, comprising:

- 5 (a) operationally linking a diverse population of first oligonucleotides having a desirable bias of random codon sequences to a first vector;
- 10 (b) operationally linking a diverse population of second oligonucleotides having a desirable bias of random codon sequences to a second vector;
- 15 (c) combining the vector products of steps (a) and (b) under conditions where said populations of first and second oligonucleotides are joined together into a population of combined vectors;
- 20 (d) introducing said population of combined vectors into a compatible host under conditions sufficient for expressing said population of random peptides;
- (e) determining the peptide which binds to said ligand binding protein;
- 25 (f) isolating the nucleic acid encoding said peptide; and
- (g) sequencing said nucleic acid.

50. The method of claim 49, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is unbiased.

51. The method of claim 49, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is diverse but biased toward a predetermined sequence.

52. The method of claim 49, wherein said first and second oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.

53. The method of claim 49, wherein steps (a) through (c) are repeated two or more times.

54. A method of constructing a diverse population of vectors containing expressible oligonucleotides having a desirable bias of random codon sequences, comprising operationally linking a diverse 5 population of oligonucleotides having a desirable bias of random codon sequences to expression elements.

55. The method of claim 54, wherein said oligonucleotides are expressible as fusion proteins on the surface of filamentous bacteriophage.

56. The method of claim 54, wherein said filamentous bacteriophage are M13.

57. The method of claim 54, wherein said fusion protein contains the product of gene VIII.

58. The method of claim 54, wherein the desirable bias of random codon sequences of said oligonucleotides is unbiased.

59. The method of claim 54, wherein the desirable bias of random codon sequences of said oligonucleotides is diverse but biased toward a predetermined sequence.

60. The method of claim 54, wherein said oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.

61. The method of claim 54, wherein said operationally linking further comprising the steps of:

5 (a) operationally linking a diverse population of first oligonucleotides having a desirable bias of random codon sequences to a first vector;

10 (b) operationally linking a diverse population of second oligonucleotides having a desirable bias of random codon sequences to a second vector; and

15 (c) combining the vector products of steps (a) and (b) under conditions where said populations of first and second oligonucleotides are joined together into a population of combined vectors.

62. The method of claim 61, wherein steps (a) through (c) are repeated two or more times.

63. A method of selecting a peptide capable of being bound by a binding protein from a population of random peptides, comprising:

5

10

- (a) operationally linking a diverse population of oligonucleotides having a desirable bias of random codon sequences to expression elements;
- (b) introducing said population of vectors into a compatible host under conditions sufficient for expressing said population of random peptides; and
- (c) determining the peptide which binds to said ligand binding protein.

64. The method of claim 63, wherein said population of random peptides are expressed as fusion proteins on the surface of filamentous bacteriophage.

65. The method of claim 63, wherein said filamentous bacteriophage are M13.

66. The method of claim 63, wherein said fusion protein contains the product of gene VIII.

67. The method of claim 63, wherein the desirable bias of random codon sequences of said oligonucleotides is unbiased.

68. The method of claim 63, wherein the desirable bias of random codon sequences of said oligonucleotides is diverse but biased toward a predetermined sequence.

69. The method of claim 63, wherein said oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.

70. The method of claim 63, wherein step (a) further comprises:

5

(a1) operationally linking a diverse population of first oligonucleotides having a desirable bias of random codon sequences to a first vector;

10

(a2) operationally linking a diverse population of second oligonucleotides having a desirable bias of random codon sequences to a second vector; and

15

(a3) combining the vector products of steps (a) and (b) under conditions where said populations of first and second oligonucleotides are joined together into a population of combined vectors.

71. The method of claim 70, wherein steps (a1) through (a3) are repeated two or more times.

72. A method of determining the nucleic acid sequence encoding a peptide capable of being bound by a ligand binding protein which is selected from a population of random peptides, comprising:

5 (a) operationally linking a diverse population of oligonucleotides having a desirable bias of random codon sequences to expression elements.

10 (b) introducing said population of vectors into a compatible host under conditions sufficient for expressing said population of random peptides;

(c) determining the peptide which binds to said ligand binding protein;

15 (d) isolating the nucleic acid encoding said peptide; and

(e) sequencing said nucleic acid.

73. The method of claim 72, wherein said population of random peptides are expressed as fusion proteins on the surface of filamentous bacteriophage.

74. The method of claim 72, wherein said filamentous bacteriophage are M13.

75. The method of claim 72, wherein said fusion protein contains the product of gene VIII.

76. The method of claim 72, wherein the desirable bias of random codon sequences of said oligonucleotides is unbiased.

77. The method of claim 72, wherein the desirable bias of random codon sequences of said oligonucleotides is diverse but biased toward a predetermined sequence.

78. The method of claim 72, wherein said oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.

79. The method of claim 72, wherein step (a) further comprises:

5

(a1) operationally linking a diverse population of first oligonucleotides having a desirable bias of random codon sequences to a first vector;

10

(a2) operationally linking a diverse population of second oligonucleotides having a desirable bias of random codon sequences to a second vector; and

15

(a3) combining the vector products of steps (a) and (b) under conditions where said populations of first and second oligonucleotides are joined together into a population of combined vectors.

80. The method of claim 78, wherein steps (a1) through (a3) are repeated two or more times.

81. A vector comprising two copies of a gene encoding a filamentous bacteriophage coat protein, both copies encoding substantially the same amino acid sequence but having different nucleotide sequences.

82. The vector of claim 81, wherein said filamentous bacteriophage is M13.

83. The vector of claim 81, wherein said gene is gene VIII.

84. The vector of claim 81, wherein said vector has substantially the sequence shown in Figure 5 (SEQ ID NO: 1).

85. A vector comprising two copies of a gene encoding a filamentous bacteriophage coat protein, one copy of said gene capable of being operationally linked to an oligonucleotide wherein said oligonucleotide can be expressed as a fusion protein on the surface of said 5 filamentous bacteriophage or as a soluble peptide.

86. The vector of claim 84, wherein said one copy of said gene is expressed on the surface of said filamentous bacteriophage.

87. The vector of claim 84, wherein said bacteriophage coat protein is M13 gene VIII.

1/16

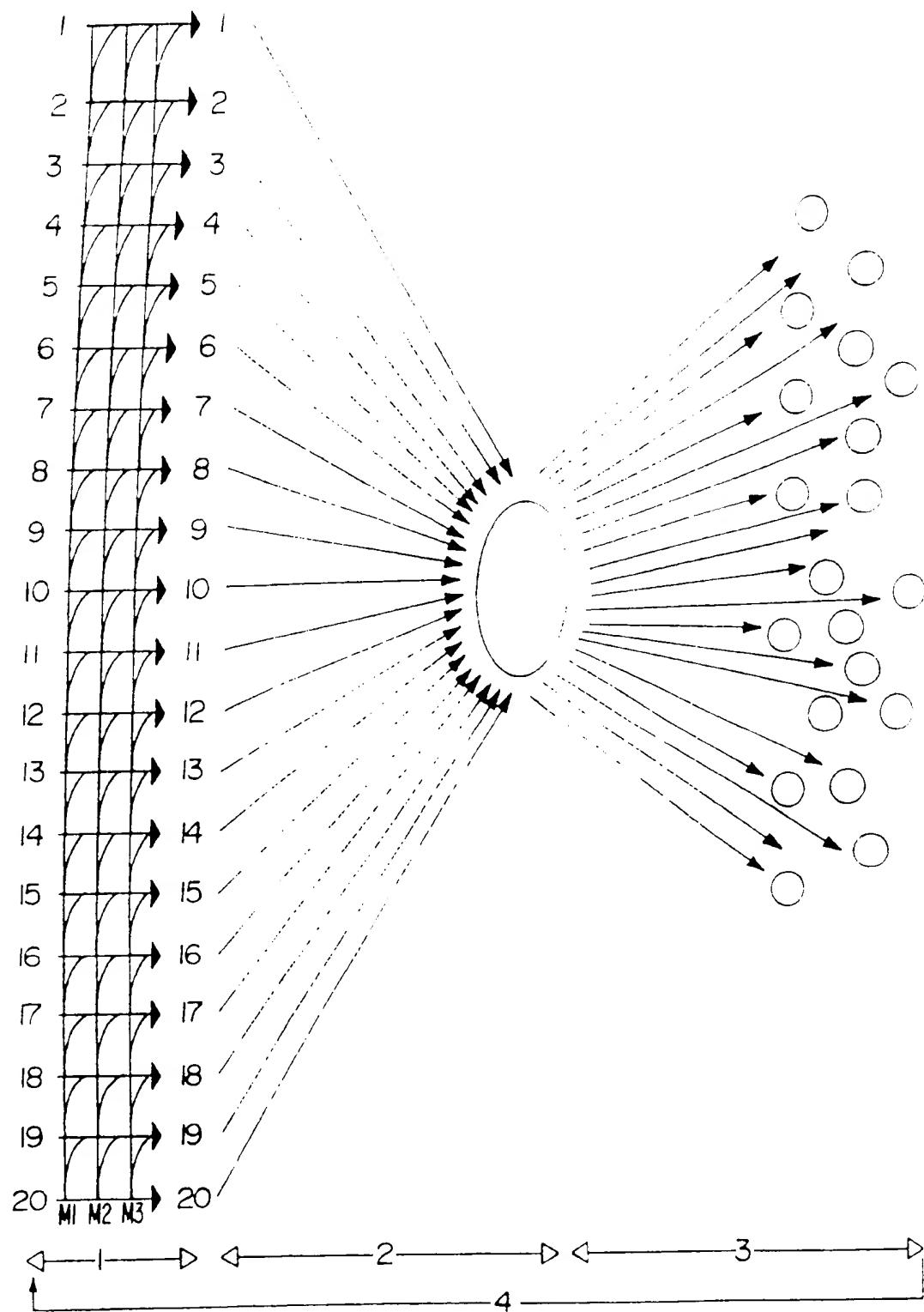


FIG. 1

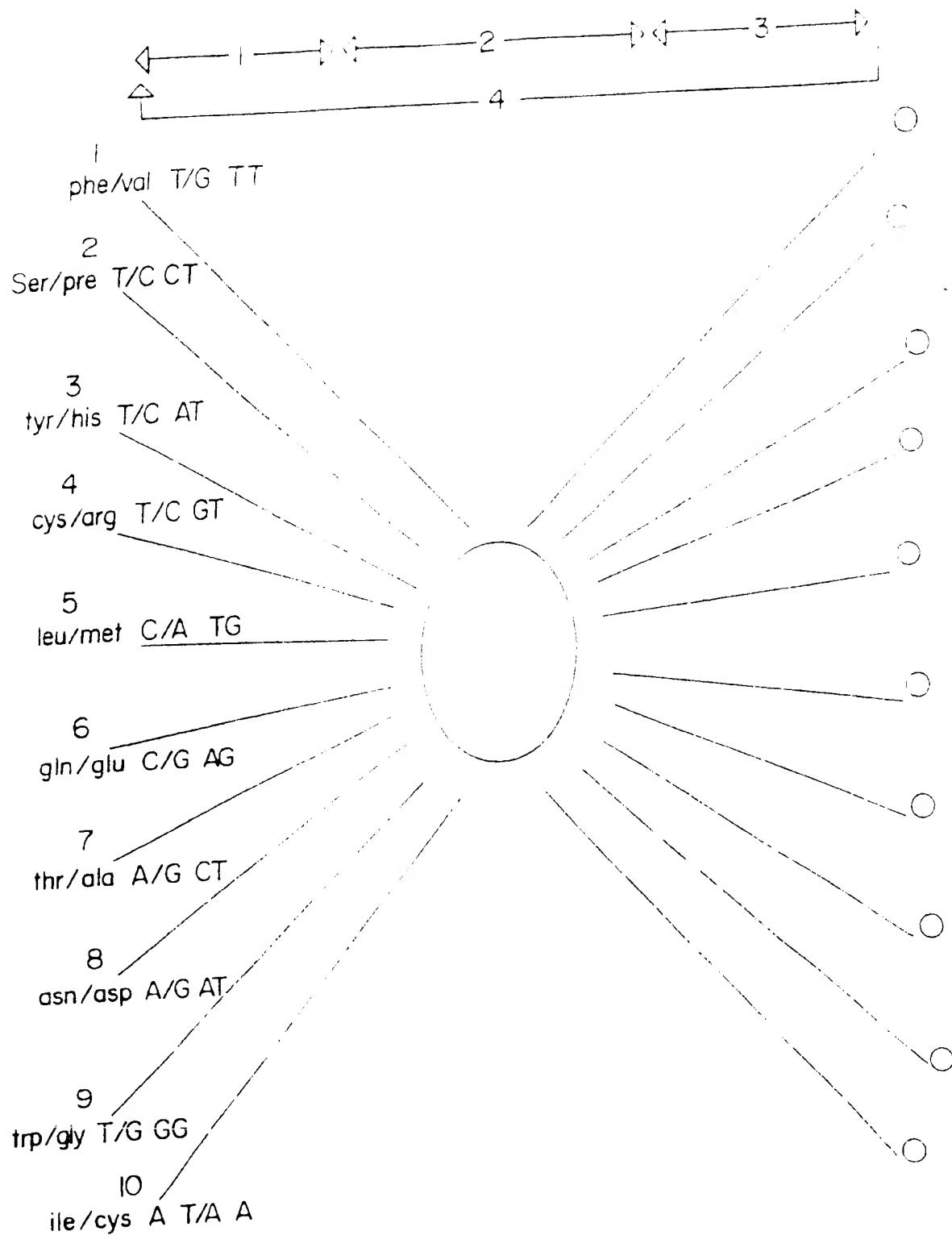


FIG. 2

3,715

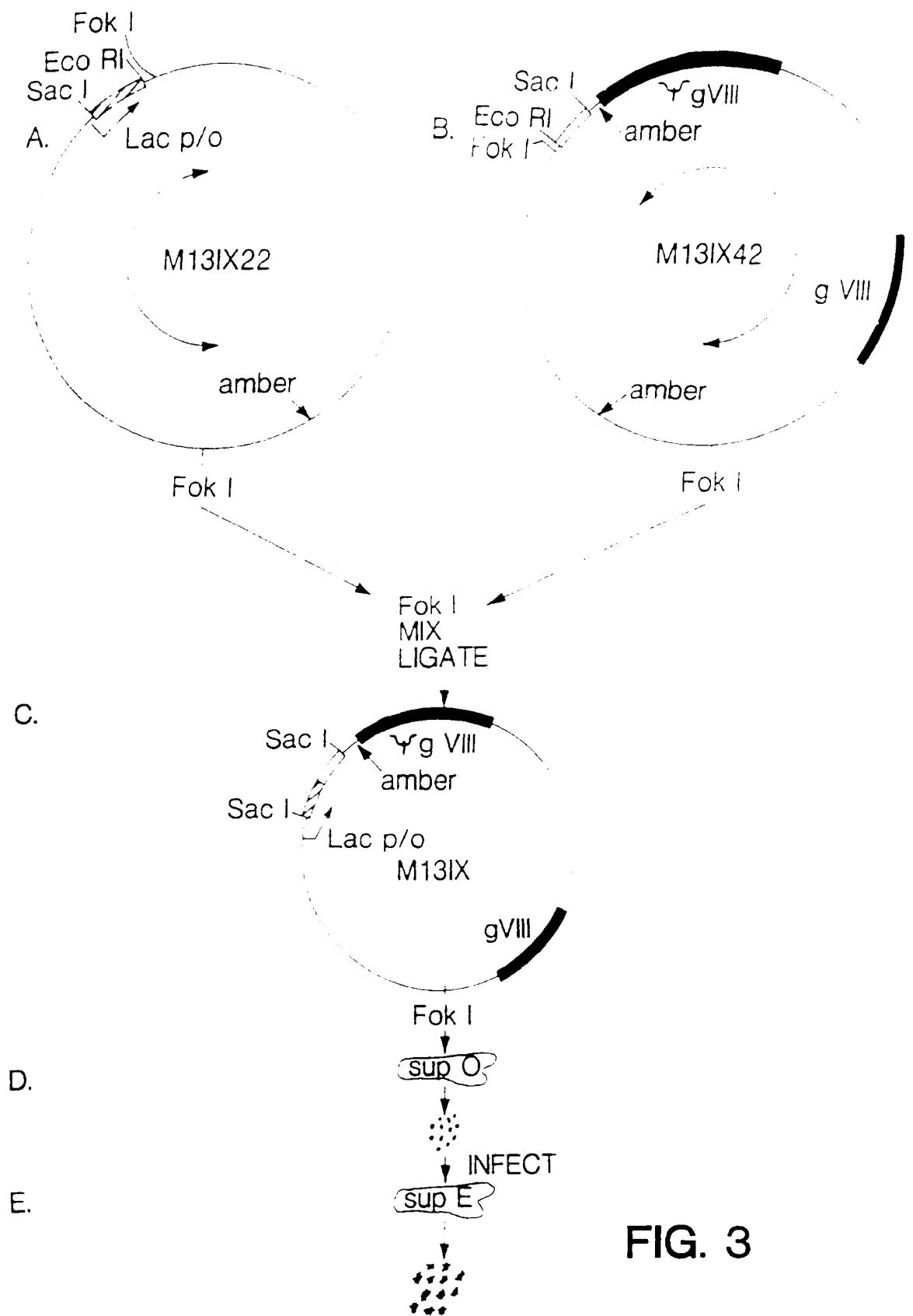
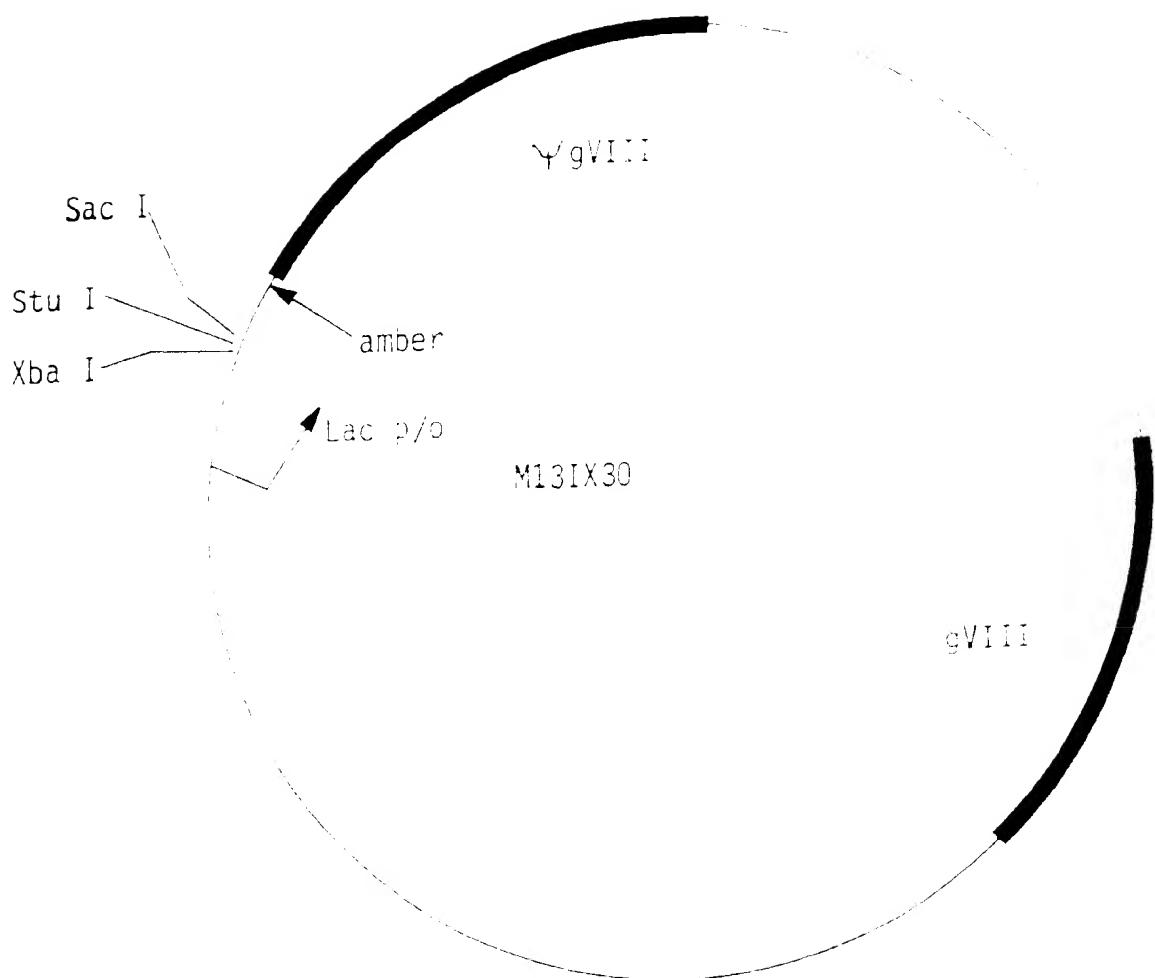


FIG. 3

4716

**FIG. 4**

5 / 15

	10	20	30	40	50	60
1	AATGCTACTA	CTATTAGTAG	AATTGATGCG	ACGTTTTGCG	CTCGCGCGCG	AAATGAAAT
61	ATAGCTAAAC	AGGTTATTGAA	CCATTGCGGA	AATGTATCT	ATGGTCAAAAC	AAATCTACT
121	CGTTGCGAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCCAGACAA	CCGTACTTTA
181	GTTGCATATT	AAAACATGT	TGAGCTCACG	SACCAAGATT	AGCAATTAAAG	CTCTAACCGA
241	TCTGCAAAAA	TGACCTCTTA	TCAAAAGGAG	CAATTAAAGG	ACCTCTCTAA	TCCTGACCTG
301	TTGGAGTTG	CTTCGGGTCT	GGTTCGCTT	GAAGCTCGAA	TTAAACGCG	ATATTGAAAG
351	TCTTCGGGC	TTCCCTCTTAA	TCTTTTGAT	GAATTCGCT	CTGCTCTGA	CTATAATAGT
421	CAGGGTAAG	ACCTGATTTT	TGATTTATGG	TCATTCTCGT	TTCTGAACT	GTTTAAGCA
481	TTTGAGGGGG	ATTCATGAA	TATTTATGAC	GATTCCGCAG	TATGGACCG	TATCCAGTCT
541	AAACATTTTA	CTATTACCCC	CTCTGGCAAA	ACCTCTTTG	CAAAGCCTC	TCGCTATTTT
601	GGTTTTTATC	GTCGTCTGGT	AAACGAGGGT	TATGATAGTG	CTGCTCTTAC	TATGCCTCGT
661	AATTCCCTTT	GGCGTTATGT	ATCTGCATTA	GTTGAATGTC	CTATTCTCAA	ATCTCAACIG
721	ATGAATCTTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTTC	CTTTTATTAA	CGTAGATTT
781	TCTTCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTTCTTA	AAATCGCATA	AGGTAATTCA
841	CAATGATTAA	AGTTGAAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCTGGTGT
901	CTCGTCAGGG	CAAGCTTAT	TCACIGAATG	AGCAGCTTG	TTACGTTGAT	TTGGGTAATG
961	AATATCCGGT	TCTTGTCAAG	ATTACTCTTG	ATGAAGGTCA	CCAGCCTAT	GCGCCTGGTC
1021	TGTACACCGT	TCATCTGTCC	TCTTCAAAG	TTGGTCAGTT	CGGTTCCCTT	ATGATTGACG
1081	GTCTGGCGCT	CGTTCCGGCT	AAGTAACATG	GAGCAGGTG	GGGATTTCGA	CACAATTTH
1141	CAGGCGATGA	TACAAATCTC	CGTTGTACTT	TGTTTCCGGC	TTGGTATAAT	GGCTGGGGGT
1201	CAAAGATGAG	TGTTTGTAGTG	TATTCTTCG	CCCTCTTCTG	TTTAGGTTG	TGCTTCTCGT
1261	GTGGCATTAC	GTATTTTACG	CGTTTAAATGG	AAACTCTTC	ATGAAAAAGT	CTTTAGTCTC
1321	CAAAGCCTCT	GTAGCCGTG	CTACCCCTCGT	TCCGATGCTG	TCTTCGCTG	CTGAGGGTGA
1381	CGATCCCGCA	AAAGCGGCT	TTAACCTCCCT	GCAGGCTCA	GGGACCGAA	ATATCGGTTA
1441	TGCGTGGGCG	ATGGTTGTG	TCATGTCGG	CGCAACTATC	GGTATCAAGC	TGTTTAAGAA
1501	ATTCACCTCG	AAAGCAAGCT	GATAAAACCGA	TACAATTAAA	GGCTCCTTTT	GGAGCCTTT
1561	TTTTTGGAGA	TTTCACAGT	AAAAAAATTA	TTATTGCAA	TTCTTTAGT	TGTTCCCTTC
1621	TATTCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTTAGCAA	AACCCCATAC	AGAAAATTCA
1681	TTTACTAACG	TCTGGAAAGA	CGACAAAAACT	TTAGATCGT	ACGCTAACTA	TGAGGGTTG
1741	CTGTGGAATG	CTACAGGCGT	TGTAGTTG	ACTGGTGAACG	AAACTCAGTG	TTACGGTACA
1801	TGGGTTCTA	TGGGCTTG	TATCCCTGAA	ATGAGGGGTG	GTGGCTCTGA	GGGTGGCGGT
1861	TCTGAGGGTG	GCGGTTCTGA	GGGTGGCGGT	ACTAACCTC	CTGAGTACGG	TGATACACCT
1921	ATTCCGGGCT	ATACTTATAT	CAACCTCTC	GAACGGCACTT	ATCCGCTGG	TACTGAGCAA
1981	AACCCCGCTA	ATCCTAATCC	TCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCATGTTT
2041	CAGAATAATA	GGTTCGAAA	TAGGCAGGGG	GCATTAACIG	TTTATACGGG	CACTGTTACT
2101	CAAGGCACTG	ACCCCGTTAA	AACTTATTAC	CAGTACACTC	CTGTATCATC	AAAAGCCATG
2161	TATGACGCTT	ACTGGAACGG	TAATTCAAGA	GAATGCGCTT	TCCATTCTGG	CTTTAATGAA
2221	GATCCATTG	TTTGTGAATA	TCAAGGCCA	TCGTCTGACC	TCGGTCAACC	TCCTGTCAAT
2281	GCTGGCGCG	GCTCTGGTGG	TGGTTCTGGT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT
2341	GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGG	GGCGGTTCCG	GTGGTGGCTC	TGGTTCCGGT
2401	GATTTGATT	ATGAAAAGAT	GGCAAACGCT	AATAAGGGGG	CTATGACCGA	AAATGCCGAT
2461	AAAAACCGCG	TACAGTCTGA	CGCTAAAGGC	AAACTTGAT	CTGTGCTAC	TGATTACGGT
2521	GCTGCTATCG	ATGGTTTCA	TGGTGAACGTT	TCCGGCCTTG	CTAATGGTAA	TGGTGCTACT
2581	GGTGATTTG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTACACCT
2641	TTAATGATA	ATTTCCTGCA	ATATTACCT	TCCCTCCCTC	AATCGTTGA	ATGTCGCCCT
2701	TTTGTCTTA	GGCCTGGTAA	ACCATATGAA	TTTCTATTG	ATTGTGACAA	AATAAACTTA
2761	TCCTGTTGGT	TCTTTGCGTT	TCTTTATAT	TTTGCACCT	TTATGTATGT	ATTTTCTACG
2821	TTTGCTAAC	TACTGCGTAA	TAAGGAGTCT	TAATCATGCC	AGTTCTTTG	GGTATTCCG
2881	TATTATTGCG	TTTCTCTGGT	TTCTCTCTGG	TAACTTTGTT	GGCGTATCTG	CTTACTTTTC
2941	TTAAAAAGGG	CTTCGGTAAAG	ATAGCTATTG	CTATTCATT	GTTTCTTGCT	CTTATTATTG
3001	GGCTTAACCTC	AATTCTTGTG	GGTTATCTCT	CTGATATTAG	CGCTCAATT	CCCTCTGACT
3061	TTGTTCAAGGG	TGTTCACTT	ATTCTCCCGT	CTAATGCGCT	TCCCTGTTT	TATGTTATT
3121	TCTCTGTAA	GGCTGCTATT	TTCATTTTG	ACGTTAAACA	AAAATCGTT	TCTTATTG
3181	ATTGGGATAA	ATAATATGGC	TGTTTATTTT	GTAACTGGCA	AATTAGGCTC	TGGAAAGACG
3241	CTCGTGTAGCG	TTGGTAAAGAT	TCAGGATAAA	ATTGTAGCTG	GGTCAAAAT	AGCAACTAA
3301	CCTGATTTAA	GGCTTCAAAA	CCTCCCGCAA	GTCGGGAGGT	TCGCTAAAC	GCCTCGCGTT
3361	CTTAGAATAC	CGGATAAGCC	TTCTATATCT	GATTTGCTTG	CTATTGGGCG	GGTAATGAT
3421	TCCTACGATG	AAAAAATAAA	CGGCTGCTT	GTTCTCGATG	AGTGGGTTAC	TGGGTTAA
3481	ACCCGTTCTT	GGAAATGATAA	GGGAAAGACAG	CCGATTATTG	ATTGGTTTAC	ACATGCTCGT
3541	AAATTAGGAT	GGGATATTAT	CTTCCTTGTG	CAGGACTTAT	CTATTGTTGA	TAAACAGGCG
3601	CGTTCTGCAT	TAGCTGAACA	TCTTGTATAT	TGTCGTGTC	GGGACAGAA	TACTTTACCT
3661	TTTGTGGTGA	CTTTATATTTC	TCTTATTACT	GGCTGAAAAA	TCCTCTGCG	TAAATTACAT
3721	TTTGGCGTGTG	TTAAATATGG	CGATTCTCAA	TTAAGCCCTA	CGGAGGCG	TGGCTTTAT
3781	ACTGGTAAGA	ATTGTATAA	CGCATATGAT	ACTAAACAGG	CTTTCTGATG	TGATTATGAT

FIG. 5-1

3841	TCCGGTGT	TT ATTCTTATT	AACGCCTTAT	GCTTACTAAA	TTATCACACG	GTCGGTATT	CAACCATT	3900
3901	AATTAGGTC	AGAAGATGAA	GCTTACTAAA	ATATAATTGA	AAAAGTTTC	ACGCGTTCTT	3960	
3961	TGTCTTGC	GA TTGGATTTC	ATCAGCAT	ACATATAGTT	ATATAACCCA	ACCTAAGCCG	4020	
4021	GAGGTAA	AGGTAGTCTC	TCAGACCTAT	GATTTGATA	AATTCACTAT	TGACTCTTCT	4080	
4081	CAGCGTCTT	ATCTAAGCTA	TCGCTATGTT	TTCAAGGATT	CTAAGGGAAA	ATTAATTAAT	4140	
4141	AGCGACGATT	TACAGAAGCA	AGGTATTCA	CTCACATATA	TTGATTTATG	TACTGTTTCC	4200	
4201	ATTAAAAAGG	TAATTCAAAT	GAAATTGTTA	AATGTAATT	ATTTTGT	CTTGATGTTT	4260	
4261	GTTCATCAT	CTTCTTTG	TCAGGTAATT	GAAATGAATA	ATTCGCCTCT	GCGCGATTTT	4320	
4321	GTAACTTGGT	ATTCAAAGCA	ATCAGGCGAA	TCGTTATTG	TTTCTCCCGA	TGTAAAAGGT	4380	
4381	ACTGTTACTG	TATATTCATC	TGACGTTAAA	CCTGAAAATC	TACGCAATT	CTTTATTCT	4440	
4441	GTTCACGTG	CTAATAATT	TGATATGGT	GGTTCATT	CTTCCATT	TTAGAAGTAT	4500	
4501	AATCCAAACA	ATCAGGATT	TATTGATGAA	TTGCATCAT	CTGATAATCA	GGAATATGAT	4560	
4561	GATAATTCCG	CTCCTTCTG	TGGTTTCTT	GTTCCGCAA	ATGATAATGT	TACTCAAAC	4620	
4621	TTTAAATTA	ATAACGTTCG	GGCAAAGGAT	TTAACACGAG	TTGTCGAATT	GTTGTAAAG	4680	
4681	TCTAACTCT	CTAAATCCTC	AAATGTATT	TCTATTGACG	GCTCTAATCT	ATTAGTTGTT	4740	
4741	AGTGCACCTA	AAGATATT	AGATAACCTT	CCTCAATT	TTTCTACTGT	TGATTGCCA	4800	
4801	ACTGACCAGA	TATTGATTGA	GGGTTTGATA	TTTGAGGTT	AGCAAGGTGA	TGCTTAGAT	4860	
4861	TTTCATTTG	CTGCTGGCTC	TCAGCGTGC	ACTGTTGCAG	GCGGTGTTAA	TACTGACCGC	4920	
4921	CTCACCTCTG	TTTATCTTC	TGCTGGTGT	TCGTCGGTAA	TTTTTAATGG	CGATGTTTTA	4980	
4981	GGGCTATCAG	TTCGCGCATT	AAAGACTAAT	AGCCATTCAA	AAATATTGTC	TGTGCCACGT	5040	
5041	ATTCTTACGC	TTCAAGGTCA	GAAGGGTCT	ATCTCTGTT	GCCAGAATGT	CCCTTTATT	5100	
5101	ACTGGTCGTG	TGACTGGTGA	ATCTGCCATT	GTAATAATC	CATTTCAGAC	GATTGAGCGT	5160	
5161	CAAAATGTAG	GTATTCCAT	GAGCCTTTT	CCTGGTGC	TGGCTGGCGG	TAATATTGTT	5220	
5221	CTGGATATT	CCAGCAAGGC	CGATAGTTG	AGTTCTTCTA	CTCAGGCAAG	TGATGTTATT	5280	
5281	ACTAATCAA	GAAGTATTGC	TACAACGGTT	AATTGCGT	ATGGACAGAC	TCTTTACTC	5340	
5341	GGTGGCCTCA	CTGATTATA	AAACACTTCT	CAAGATTCTG	GCGTACCGTT	CCTGTCTAAA	5400	
5401	ATCCCTTAA	TCGGCCTCT	GTTTAGCTC	CGCTCTGATT	CCAACGAGGA	AAGCACGTTA	5460	
5461	TACGTGCTG	TCAAAGCAAC	CATAGTACGC	GCCCTGTAGC	GGCGCATTAA	GCGCGGCCGG	5520	
5521	TGTGGTGGT	ACGGCGAGCG	TGACCCCTAC	ACTTGCAGC	GCCCTAGCGC	CCGCTCCCTT	5580	
5581	CGCTTCTTC	CCTTCTTT	TCGCCACGT	CGCCGGCTT	CCCCGTCAG	CTCTAAATCG	5640	
5641	GGGGCTCCT	TTAGGGTTC	GATTTAGTGC	TTTACGGCAC	CTCGACCCCA	AAAAACTTGA	5700	
5701	TTGGGTGAT	GGTCACGTA	GTGGGCCATC	GCCCTGATAG	ACGGTTTTTC	GCCCTTTGAC	5760	
5761	GTGGAGTCC	ACGTTCTTA	ATAGTGGACT	CTTGTCCAA	ACTGGAACAA	CACTCAACCC	5820	
5821	TATCTCGGGC	TATTCTTTG	ATTTATAAGG	GATTTCGCC	ATTCGGAAC	CACCATAAA	5880	
5881	CAGGATTTG	GCCTGCTGG	GCAAACCAAGC	GTGGACCGCT	TGCTGCAACT	CTCTCAGGGC	5940	
5941	CAGGCGGTGA	AGGGCAATCA	GCTTGGCCC	GTCTCGCTGG	TGAAAAGAAA	AACCACCCCTG	6000	
6001	GCGCCCAATA	CGCAAACCGC	CTCTCCCCGC	GCCTTGGCCG	ATTCAATT	GCAGCTGGCA	6060	
6061	CGACAGGTT	CCCGACTGG	AAGCGGGCAG	TGAGCGCAAC	GCAATTAAATG	TGAGTTAGCT	6120	
6121	CACTCATTAG	GCACCCCAGG	CTTTACACTT	TATGCTTCCG	GCTCGTATGT	TGTGTGAAT	6180	
6181	TGTGAGCGGA	TAACATTTC	ACACAGGAAA	CAGCTATGAC	CAGGATGTC	GAATTGCGAG	6240	
6241	GTAGGAGAGC	TCGGCGGATC	CTAGGCTGAA	GGCGATGACC	CTGCTAAGGC	TGCAATTAAAT	6300	
6301	AGTTTACAGG	CAAGTGTAC	TGAGTACATT	GGCTACGCTT	GGGCTATGGT	AGTAGTTATA	6360	
6361	GTGGTGTCA	CCATAGGGAT	TAATTATTTC	AAAAAGTTTA	CGAGCAAGGC	TTCTTAACCA	6420	
6421	GCTGGGTAA	TAGCGAAGAG	GCCCCGACCG	ATGCCCTTC	CCAACGATTG	CGCAGCCTGA	6480	
6481	ATGGCGAATG	GCGCTTTGCC	TGGTTCCGG	CACCAGAACG	GGTGCCTGGAA	AGCTGGCTGG	6540	
6541	AGTGCATCT	TCCGTAGG	GATACGGTCG	TCGTCCTTC	AAACTGGCAG	ATGCACGGTT	6600	
6601	ACGATGCGCC	CATCTACACC	AACGTAACCT	ATCCCATTAC	GGTCAATCCG	CCGTTTGTT	6660	
6661	CCACGGAGAA	TCCGACGGGT	TGTTACTCGC	TCACATTAA	TGTTGATGAA	AGCTGGCTAC	6720	
6721	AGGAAGGCCA	GACCGAATT	ATTTTTGATG	GGCTTCCCTAT	TGGTTAAAAA	ATGAGCTGAT	6780	
6781	TTAACAAAAA	TTAACGCGA	ATTTAACAA	AAATTAACG	TTTACAATT	AAATATTGTC	6840	
6841	TTATACAATC	TTCTGT	TGGGGTTT	CTGATTATCA	ACCGGGTAC	ATATGATTGA	6900	
6901	CATGCTAGTT	TTACGATTAC	CGTCATCGA	TTCTCTTGT	TGCTCCAGAC	TCTCAGGGCAA	6960	
6961	TGACCTGATA	GCCTTGTAG	ATCTCTAAA	AATAGCTACC	CTCTCAGGCA	TTAATTATAC	7020	
7021	AGCTAGAACG	GTTGAATATC	ATATTGATGG	TGATTGACT	GTCTCAGGCCC	TTTCTCACCC	7080	
7081	TTTTGAATCT	TTACCTACAC	ATTACTCAGG	CATTGACTT	AAAATATATG	AGGGTTCTAA	7140	
7141	AAATTTTAT	CCTTGC	AAATAAGGC	TTCTCCCGCA	AAAGTATTAC	AGGGTCATAA	7200	
7201	TGTTTGGT	ACAACCGATT	TAGCTTATG	CTCTGAGGCT	TTATTGCTTA	TTTTGCTAA	7260	
7261	TTCTTGCCT	TGCCTGTATG	ATTTATTGGA	CGTT	40	50	60	

FIG. 5-2

SUBSTITUTE SHEET

7/15

	10	20	30	40	50	60
1	AATGCTACT	CTATTAGTAG	AATTGATGCG	ACGTTTCAG	CTCGCGCGCG	AAATGAAAT
61	ATAGCTAAC	AGGTTATTGA	CCATTGGCGA	AACTGATCTA	ATGGTCAAA	TAAATCTACT
121	CGTCGAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCCAGACA	CCGTACTTTA
181	GTTGCATATT	TAAAACATGT	TGAGCTACAG	CACCAAGATT	AGCAATTAAG	CTCTAAGCCA
241	TCTGCAAAAA	TGACCTCTTA	TCAAAAGGAG	CAATTAAAGG	TACTCTCTAA	TCCTGACCTG
301	TTGGAGTTG	CTTCGGTCT	GGTCGCTT	GAAGCTCGAA	TTAAAACGCG	ATATTTGAAG
361	TCTTCGGGC	TTCTCTTAA	TCTTTTGAT	GCAATCCGCT	TTGCTTCTGA	CTATAATAGT
421	CAGGGTAAAG	ACCTGATTT	TGATTTATGG	TCATTCTCGT	TTTCTGAACT	GTTAAAGCA
481	TTTGAGGGGG	ATTCAATGAA	TATTATGAC	GATTCCGCG	TATTGGACGC	TATCCAGTCT
541	AAACATTTTA	CTATTACCC	CTCTGGCAAA	ACTTCTTTG	CAAAGCCTC	TCGCTATTT
601	GGTTTTATC	GTCGCTGGT	AAACGAGGGT	TATGATAGTG	TTGCTCTTAC	TATGCCCTCGT
661	AATTCCCTTT	GGCGTTATGT	ATCTGCATTA	GTGAATGTTG	GTATTCTAA	ATCTCAACTG
721	ATGAATCTT	CTACCTGTAA	TAATGTTGTT	CGGTTAGTTC	GTTTTATTA	CGTAGATTT
781	TCTTCCAAAC	GTCCTGACTG	GTATAATGAG	CCAGTCTTA	AAATCGCATA	AGGTAATTCA
841	CAATGATTAA	AGTTGAATT	AAACCATCTC	AAGCCCATT	TACTACTCGT	TCTGGTGTGTT
901	CTCGTCAGGG	CAAGCCTT	TCACGTGATG	AGCAGCTT	TTACGTTGAT	TTGGGTAATG
961	AATATCCGGT	TCTTGTCAAG	ATTACTCTTG	ATGAAGGTC	GCCAGCCTAT	GCGCTGGTC
1021	TGTACACCGT	TCATCTGTCC	TCTTCAAAG	TTGGTCAGTT	CGGTTCCCTT	ATGATTGACC
1081	GTCTGCGCCT	CGTTCCGGCT	AAGTAACATG	GAGCAGGTCG	CGGATTTCGA	CACAATTAT
1141	CAGGGCATGA	TACAAATCTC	CGTTGACTT	GTGTTCGCGC	TTGGTATAAT	CGCTGGGGGT
1201	CAAAGATGAG	TGTTTAGTG	TATTCTTCG	CCTCTTCGT	TTTAGCTTGG	TGCCCTCGTA
1261	GTGGCATTAC	GTATTTTAC	CGTTTAATGG	AAACTCCCTC	ATGAAAAAGT	CTTTAGTCCT
1321	CAAAGCCTCT	GTAGCCGTTG	CTACCCCTCGT	TCCGATGCTG	TCTTCGCTG	CTGAGGGTGA
1381	CGATCCCGCA	AAAGCGGCCT	TTAACCTCCCT	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA
1441	TGCGTGGGCG	ATGGTTGTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTTAAGAA
1501	ATTCACCTCG	AAAGCAAGT	GATAAACCGA	TACAATTAAA	GGCTCCCTTT	GGAGCCTTT
1561	TTTTGGAGA	TTTCAACGT	GAAAAAATT	TTATTGCAA	TTCCCTTAAAGT	TGTTCCCTTTC
1621	TATTCCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTAGCCTA	AAACCCATAC	AGAAAATTCA
1681	TTTACTAACG	TCTGGAAAGA	CGACAAAATT	TTAGATCGT	ACGCTAACTA	TGAGGGTTGT
1741	CTGTGGAATG	CTACAGGCGT	TGTAGTTGT	ACTGGTGAACG	AAACCTCAGTG	TTACGGTACA
1801	TGGGTTCTA	TTGGGCTTGC	TATCCCTGAA	AATGAGGGGTG	GTGGCTCTGA	GGGTGGCGGT
1861	TCTGAGGGGTG	GCGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGAGTACGG	TGATACACCT
1921	ATTCCGGGCT	ATACTTATAT	CAACCCCTCTC	GACGGCACCT	ATCCGCCTGG	TACTGAGCAA
1981	AACCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCATGTTT
2041	CAGAATAATA	GTTCGAA	TAGGCAGGGG	GCATTAAC	TTTATACGGG	CACTGTTACT
2101	CAAGGCACTG	ACCCCGTTAA	AACTTATTAC	CAGTACACTC	CTGTATCATC	AAAAGCCATG
2161	TATGACGCTT	ACTGGAACGG	TAATTCAGA	GACTGCGCTT	TCCATTCTGG	CTTTAATGAA
2221	GATCATTCTG	TTTGTGAATA	TCAAGGCCAA	TCGTCTGACC	TGCCCTCAACC	TCCTGTCAAT
2281	GCTGGCGCG	GCTCTGGTGG	TGGTTCTGGT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT
2341	GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGG	GGCGGTTCCG	GTGGTGGCTC	TGGTTCCGGT
2401	GATTTTGATT	ATGAAAAGAT	GGCAAACGCT	AATAGGGGG	CTATGACCGA	AAATGCCGAT
2461	GAAAACCGCG	TACAGTCTGA	CGCTAAAGGC	AAACTTGTATT	CTGTGCTAC	TGATTACGGT
2521	GCTGCTATCG	ATGGTTTCAT	TGGTGACGTT	TCCGGCCTTG	CTAATGGTAA	TGGTGCTACT
2581	GGTGATTTG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTACACT
2641	TTAATGAATA	ATTTCCGTC	ATATTTCACCT	TCCCTCCCTC	AATCGGTTGA	ATGTGCGCC
2701	TTTGTCTTTA	GCGCTGGTAA	ACCATATGAA	TTTCTATTG	ATTGTGACAA	AATAAAACTA
2761	TTCCGTGGTG	TCTTTCGTT	TCTTTATAT	GTTGCCACCT	TTATGTATGT	ATTTTCTACG
2821	TTTGCTAAC	TACTGCGTA	TAAGGAGTCT	TAATCATGCC	AGTTCTTTG	GGTATTCCCGT
2881	TATTATTGCG	TTTCCTCGGT	TTCCTCTGG	TAACTTTGTT	CGGCTATCTG	CTTACTTTTC
2941	TTAAAAAGGG	CTTCGGTAAG	ATAGCTATTG	CTATTTCTATT	GTTTCTTGCT	CTTATTATTG
3001	GGCTTAAC	AACTCTTG	GGTTATCTC	CTGATATTAG	CGCTCAATT	CCCTCTGACT
3061	TTGTTCAAGGG	TGTTCACTT	ATTCTCCCGT	GTAATGCGCT	TCCCTGTTT	TATGTTATT
3121	TCTCTGTAAA	GGCTGCTATT	TTCATTTTG	ACGTTAAACA	AAAAATCGTT	TCTTATTGTTG
3181	ATTGGGATAA	ATAATATGGC	TGTTTATTT	GTAACTGGCA	AATTAGGCTC	TGGAAAGACG
3241	CTCGTTAGCG	TTGGTAAGAT	TTAGGATAAA	ATTTGAGCTG	GGTGCACAAAT	AGCAACTAAT
3301	CTTGATTAA	GGCTTCAAA	CCTCCCGCAA	GTCGGGAGGT	TCGCTAAAC	GCCTCGCGTT
3361	CTTAGAATAC	CGGATAAGCC	TTCTATATCT	GATTTGCTTG	CTATTGGCG	CGGTAATGAT
3421	TCCTACGATG	AAAATAAAAA	CGGCTGCTT	GTTCTCGATG	AGTGCCTGAC	TTGGTTTAAT
3481	ACCCGTTCTT	GGAATGATAA	GGAAAGACAG	CCGATTATTG	ATTGGTTCT	ACATGCTCGT
3541	AAATTAGGAT	GGGATATTAT	CTTCCTTGTT	CAGGACTTAT	CTATTGTTGA	AAAACAGGCG
3601	CGTTCTGCAT	TAGCTGAACA	TGTTGTTAT	TGTCGTCGTC	TGGACAGAA	TACTTTACCT
3661	TTTGTGGTGA	CTTTATATT	TCTTATTACT	GGCTCGAAA	TGCCTCTGCC	TAAATTACAT
3721	GTTGGCGTTG	TTAAATATGG	CGATTCTCAA	TIAAGCCCTA	CTGTGAGCG	TTGGCTTTAT
3781	ACTGGTAAGA	TTTGTATAA	CGCATATGAT	ACTAAACAGG	CTTTTCTAG	TAATTATGAT
						3840

FIG. 6-1

3841	TCCGGTGT	TTT	ATTCTTATT	AACGCC	TTATCACACG	GTGGTATIT	CAAACCATTA	3900
3901	AATTAGGT	C	AGAAGATGAA	ATTAAC	ATATATTGAA	AAGAGTTTC	TCGGTTC	3950
3961	TGTCTTGC	G	TTGGATTG	ATCAGC	ACATATAGT	ATATAACCCA	ACCTAAGCCG	4020
4021	GAGGTTAAA	A	AGGTAGT	TCAGAC	GATTTGATA	AATTCACTAT	TGACTCTTCT	4080
4081	CAGCGTCT	A	ATCTAACG	TCGCTATG	TTCAAGGATT	CTAAGGGAAA	ATTAATTAA	4140
4141	AGCGACG	T	TACAGAAC	AGTTATT	CTCACATATA	TTGATTATG	TACTGTTTCC	4200
4201	ATAAAAAAAG	T	GTAATT	TGAAATTG	AAATGTAATT	AATTGTT	TCTTGATG	4260
4261	TGTTTCATCA	T	TCTTCTT	CTCAGG	TGAAATTG	AATTGTT	TGCGCGATT	4320
4321	TGTAAC	T	TATTCAAA	AATCAGG	ATCCGTT	GTTCCTCCG	ATGAAAAGG	4380
4381	TACTGTTACT	G	GTAATT	CTGACG	ACCTGAAA	CTACGCAATT	TCTTTATTTC	4440
4441	TGTTTACGT	G	GCTAATA	TTGATATG	TTGCTCAATT	TCTGATAAATC	AGGAATATGA	4500
4501	TAATCCAAAC	A	AATCAGG	ATATTGATG	ATTGCCATCA	AATGATAATG	TTACTCAAAC	4560
4561	TGATAATTCC	G	GCTCCTCTG	GTGGTTCT	TGTTCCGAA	GTGTCGAAT	TGTTGTAAA	4620
4621	TTTAAATT	A	AATAAC	GGGCAAAGG	TTTAATACGA	TATTAGTTGT	TATTAGTTGT	4680
4681	GCTTAATACT	T	TCTAAATCCT	CAAATG	ATCTATTGAC	GGCTCTAAC	TTGATTTGCC	4740
4741	TAGTGCACCT	A	AAAGATATT	TGATAAC	TCCTCAATT	CTTCTACTG	ATGCTT	4800
4801	AACTGACCAAG	A	ATATTGATTG	AGGGTTG	ATTGAGGTT	CAGCAAGGTG	ATACTGACCG	4860
4861	TTTTCAATT	G	GCTGCTGG	CTCAGC	CACTGTTGCA	GGCGGTGTTA	GCGATGTTT	4920
4921	CCTCACCTCT	G	GT	CTGCTGG	CTGCTGGTGG	ATTTTAAATG	CTGTGCCACG	4980
4981	AGGGCTATCA	G	GTTCGCG	TAAAGACTAA	TAGCCATTCA	AAATATTGT	TCCCTTTTAT	5040
5041	TATTCTTACG	C	CTTTCAGG	AGAAGGGTTC	TATCTCTGTT	GGCCAGAATG	CGATTGAGCG	5100
5101	TACTGGTCG	T	GTGACTGG	AATCTGCCA	TGAAATAAT	CCATTTCAGA	GTAAATTG	5160
5161	TCAAAATGTA	G	GTATT	TGAGCG	TCCGTG	ATGGCTGGCG	TGATGTTAT	5220
5221	TCTGGATATT	A	ACCAGCAAGG	CCGATAG	ACTCAGGCA	CTCTTTACT	5280	
5281	TACTAATCAA	A	AGAAGTATTG	CTACAA	TAATTGCGT	GATGGACAGA	CTCTGTCTAA	5340
5341	CGGTGGCCTC	T	ACTGATT	AAAACACTT	TCAAGATTCT	GGCGTACCGT	AAAGCACGTT	5400
5401	AATCCCTTA	A	ATCGGC	TGTTAG	CGGCTCTGAT	TCCAACGAGG	AAAGCAGG	5460
5461	ATACGTGCTC	G	GTCAAAG	CCATAG	CGCCCTGTAG	GGCGCATT	AGCGCGGCG	5520
5521	GTGTGGTGG	T	TACGCG	GTGACCG	CACTTGCAG	GGCCCTAGCG	CCCCTCCTT	5580
5581	TCGTTTCTT	C	CCCTTCTT	CTGCCCACG	TGCGCCGCTT	TCCCCGTCAA	GCTCTAAATC	5640
5641	GGGGGCTCCC	T	TTAGGGT	CGATTTAG	CTTACGGCA	CCTCGACCCCC	AAAAAAACTTG	5700
5701	ATTTGGGTGA	T	TGGT	AGTGGCC	CGCCCTGATA	GACGGTTTTT	CGCCCTTTGA	5760
5761	CGTTGGAGTC	C	ACGTT	AATAGTGG	TCTTGT	ACTGGAACA	CCACCATCAA	5820
5821	CTATCTCGGG	T	CTATT	GATT	GGATT	GATTTGGAA	TCTCTCAGGG	5880
5881	ACAGGATTG	T	CGCCTG	GGCAAAC	GGTGGACCG	TGCTGCAAC	5940	
5941	CCAGGCGGT	A	AAGGGCA	AGCTG	CGTGTG	GTGAAAAGAA	AAACCAACCT	6000
6001	GGCGCCCAAT	A	ACGCAAAC	CCTCT	CGCCTG	GATTCAATTAA	TGCAAGCTGGC	6060
6061	ACGACAGGTT	T	TCGGACT	AAAGCGGG	CGCCTG	GTGAGCGCAA	6120	
6121	TCACTCATTA	G	GGCAC	GCTTAC	TTATGCTTCC	CGCAATTAA	GTGAGTTAGC	6180
6181	TTGTGAGCGG	A	ATAAC	CACAGC	GGAGACAGTC	GGCTGTATG	TTGTGTTGGAA	6240
6241	TACGGCAGCC	G	GCTGATT	TATTACT	ATAATGAAAT	GGCTGTACCC	AACTTAAATCG	6300
6301	GACCCAGACT	C	CCAGAATT	ATCCGA	TTATGCTTCC	GGCTGTATG	TAAGCTTGGC	6360
6361	ACTGGCCGTC	G	TTTACAAC	GTCGT	GGAAAACCC	GGCTGTACCC	AACTTAAATCG	6420
6421	CCTTGCAGCA	C	ACCCCC	TCGCCAG	GCGTAATAGC	GAAGAGGCC	GCACCGATCG	6480
6481	CCCTTCCAA	A	CA	GGCTGA	TTTGCTGGT	TTTGCTGGT	TTCCGGCACC	6540
6541	AGAAGCGGT	C	GGGAAAG	GGCTGG	CGATCTTCT	GAGGCCGATA	CGGTGTCGT	6600
6601	CCCCTAAC	T	TGGCAGATG	ACGGTT	TGCGCCC	TACACCAACG	TAACCTATCC	6660
6661	CATTACGGTC	A	ATCCG	TTGTTCCAC	GGAGAATCCG	TCGGGTTGTT	ACTCGCTCAC	6720
6721	ATTTAATGTT	G	GATGAAAG	GGCTAC	GGAGCAGACG	CGAATTATT	TTGATGGCGT	6780
6781	TCCTATTG	T	AAAAAAATG	GCTGATT	CAAAAATTAA	ACGCGAATT	TAACAAAATA	6840
6841	TTAACGTTA	C	AAATTAAAT	ATTGCTT	ACAATCTTCC	TGTTTTTGGG	GCTTTCTG	6900
6901	TTATCAACCG	G	GGGTAC	GATTGAC	GATTACCGT	CATCGATTCT	6960	
6961	CTTGTG	C	CCAGACT	AGGCAAT	TTGTAGATCT	CTAAAAATA	7020	
7021	GCTACCTCT	C	CCGGCATT	TTTATCAG	AGAACGGTTG	AATATCATAT	TGATGGTGA	7080
7081	TTGACTGTCT	C	CCGGC	GAATCTTAC	CTACACATTA	CTCAGGCATT	7140	
7141	GCATTTAAA	T	TATATGAGGG	TTCTAAAAT	GGGTTGAAAT	AAAGGCTTCT	7200	
7201	CCCGAA	A	TATTACAGGG	TCATAATG	TTTGGTACAA	CCGATTAGC	TTATGCTCT	7260
7261	GAGGTTTAT	T	TGCTTAATT	TGCTAATT	TGCTAATT	TGTATGATT	ATTGGACGT	7320
		10	20	30	40	50	60	

FIG. 6-2

9 / 16

	10	20	30	40	50	60
1	AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTTCAG	CTCGCGCCCC	AAATGAAAAT
61	ATAGCTAAAC	AGGTTATTGA	CCATTGCGA	AATGTATCTA	ATGGTCAAAC	TAAATCTACT
121	CGTTCGCAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCCAGACA	CCGTACTTTA
181	GTTGCATATT	TAACCATGTT	TGAGCTACAG	CACCAAGATT	AGCAATTAAAG	CTCTAAGGCC
241	TCTGCAAAAA	TGACCTCTTA	TCAAAAGGAG	CAATTAAAGG	TACTCTCTAA	TCCTGACCTG
301	TTGGAGTTG	CTTCCGGTCT	GGTTGCTTT	GAAGCTCGAA	TTAAAACGCG	ATATTGAAAS
361	TCTTCGGGC	TTCTCTTAA	TCTTTTGAT	GCAATCCGCT	TTGCTTCTGA	CTATAATAGT
421	CAGGGTAAAG	ACCTGATTT	TGATTATGG	TCATTCTCGT	TTTCTGAACT	GTTTAAAGCA
481	TTTGGAGGGGG	ATTCAATGAA	TATTATGAC	GATTCCGCAG	TATTGGACGC	TATCCAGTCT
541	AAACATTTA	CTATTACCCC	CTCTGGCAAA	ACTTCTTTG	CAAAAGCCTC	TCGCTATTT
601	GGTTTTTATC	GTCGTCTGGT	AAACGAGGGT	TATGATAGTG	TTGCTCTTAC	TATGCCTCGT
661	AATTCCCTTT	GGCGTTATGT	ATCTGCATTA	GTTGAATGTG	GTATTCCTAA	ATCTCAACTG
721	ATGAATCTTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTTC	GTTTATTAA	CGTAGATTTC
781	TCTTCCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTTCTTA	AAATCGCATA	AGGTAATTCA
841	CAATGATTAA	AGTTGAAATT	AAACCATCTC	AAGGCCAAT	TACTACTCGT	TCTGGTGTG
901	CTCGTCAGGG	CAAGCCTTAT	TCACTGAATG	AGCAGCTT	TTACGTTGAT	TTGGGTAATG
961	AATATCCGGT	TCTTGTCAAG	ATTACTCTTG	ATGAAGGTCA	GCCAGCCTAT	GCGCCTGGTC
1021	TGTACACCGT	TCATCTGTCC	TCTTCAAAG	TTGGTCAGGT	CGGTTCCCTT	ATGATTGACC
1081	GTCTCGCCT	CGTTCCGGT	AAGTAACATG	GAGCAGGTG	CGGATTTCGA	CACAATTAT
1141	CAGGCGATGA	TACAAATCTC	CGTTGACTT	TGTTTCGCGC	TTGGTATAAT	CGCTGGGGGT
1201	CAAAGATGAG	TGTTTTAGTG	TATTCTTCG	CCTCTTTCTG	TTTAGGTTGG	TGCCTTCGTA
1261	GTGGCATTAC	GTATTTTAC	CGTTTAATGG	AAACTTCCTC	ATGAAAAAAGT	CTTTAGTCT
1321	CAAAGCCTCT	GTAGCCGTTG	CTACCCCTCGT	TCCGATGCTG	TCTTCGCTG	CTGAGGGTGA
1381	CGATCCCGCA	AAAGCGGCCT	TTAACCTCCCT	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA
1441	TGCGTGGGCG	ATGGTTGTTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTTAAGAA
1501	ATTCACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	GGCTCCCTTT	GGAGCCTTT
1561	TTTTGGAGA	TTTCAACGT	GAAAAAATTAA	TTATTCGCAA	TTCCCTTAGT	TGTTCCCTTC
1621	TATTCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTTAGCAA	AACCCCATAC	AGAAAAATTCA
1681	TTTACTAACG	TCTGGAAAGA	CGACAAAATC	TTAGATCGT	ACGCTAACTA	TGAGGGTTGT
1741	CTGTGGAATG	CTACAGGCGT	TGTTAGTTGT	ACTGGTGACG	AAACTCAGTG	TTACGGTACA
1801	TGGGTTCCCTA	TGGGCTTCTG	TATCCCTGAA	ATGGGGGTG	GTGGCTCTGA	GGGTGGCGGT
1861	TCTGAGGGGT	GCGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGAGTACGG	TGATACACCT
1921	ATTCCGGGCT	ATACTTATAT	CAACCCCTCTC	GACGGCACTT	ATCCGCCCTGG	TACTGAGCAA
1981	AACCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCATGTTT
2041	CAGAATAATA	GGTTCGAAA	TAGGCAGGGG	GCATTAACGT	TTTATACGGG	CACTGTTACT
2101	CAAGGCACTG	ACCCCGTTAA	AACTTATTAC	CACTACACTC	CTGTATCATC	AAAAGCCATG
2161	TATGACGCTT	ACTGGAACGG	TAATTTCAGA	GACTGCGCTT	TCCATTCTGG	CTTTAATGAA
2221	GATCCATTG	TTTGTGAATA	TCAAGGCCAA	TCGTCTGACC	TGCCTCAACC	TCCTGTCAAT
2281	GCTGGCGGCG	GCTCTGGTGG	TGGTTCTGGT	GGCGGCTCTG	AGGGGGGTG	CTCTGAGGGT
2341	GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGG	GGCGGTTCCG	GTGGGGCTC	TGGTTCCGGT
2401	GATTTGATT	ATGAAAAGAT	GGCAAACGCT	AATAAGGGGG	CTATGACCGA	AAATGCCGAT
2461	AAAAACCGCG	TACAGTCTGA	CGCTAAAGGC	AAACTTGT	CTGTCGCTAC	TGATTACGGT
2521	GCTGCTATCG	ATGGTTTCAT	TGGTGAACGT	TCCGGCTT	CTAATGGTAA	TGGTGCTACT
2581	GGTGAATTG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTACACT
2641	TTAATGATA	ATTTCCGTCA	ATATTTCACCT	TCCCTCCCTC	AATCGTTGA	ATGTGCCCT
2701	TTTGTCTTTA	GCGCTGGTAA	ACCATATGAA	TTTCTATTG	ATTGTGACAA	AATAAAACTA
2761	TTCCGTGGTG	TCTTTGCGTT	TCTTTTATAT	GTTGCCACCT	TTATGTATGT	ATTTCTACG
2821	TTTGCTAAC	TACTGCGTA	TAAGGAGTCT	TAATCATGCC	AGTTCTTTG	GGTATTCCGT
2881	TATTATTGCG	TTTCTCTGG	TAACTTTGTG	CGGCTATCTG	CTTACTTTT	2940
2941	TTAAAAAGGG	CTTCGGTAAG	ATAGCTATTG	CTATTCATT	GTTTCTTGCT	CTTATTATTG
3001	GGCTTAACTC	AATTCTTG	GGTTATCTCT	CTGATATTAG	CGCTCAATT	CCCTCTGACT
3061	TTGTTCAAGGG	TGTTCAGTAA	ATTCTCCCGT	CTAATGCCGT	TCCCTGTTT	TATGTTATTG
3121	TCTCTGTAAA	GGCTGCTATT	TTCATTTTG	ACGTTAAACA	AAAAATCGTT	TCTTATTGTTG
3181	ATTGGGATAA	ATAATATGGC	TGTTTATTTT	GTAACTGGCA	AATTAGGCTC	TGGAAAAGACG
3241	CTCGTTAGCG	TTGGTAAGAT	TCAGGATAAA	ATTGTAGCTG	GGTGC	AGCAACTAT
3301	CTTGATTAA	GGCTTCAAA	CCTCCCGCAA	GTGGGAGGT	TCGCTAAAAC	GCCTCGCGTT
3361	CTTAGAATAC	CGGATAAGCC	TTCTATATCT	GATTTGCTT	CTATTGGCG	CGGTAATGAT
3421	TCCTACGATG	AAAATAAAAAA	CGGCTTGCCT	GTTCTCGATG	AGTGC	TTGGTTTAAAT
3481	ACCCGTTCTT	GGAATGATAA	GGAAAGACAG	CGGATTATTG	ATTGTTTCT	ACATGCTCGT
3541	AAATTAGGAT	GGGATATTAT	TTTTCTTGTT	CAGGACTTAT	CTATTGTTGA	TAAACAGGGC
3601	CGTTCTGCAT	TAGCTGAACA	TGTTGTTTAT	TGTCGTCGTC	TGGACAGAAAT	TACTTTACCT
3661	TTTGTGGTAA	CTTTATATTTC	TCTTATTACT	GGCTCGAAA	TGCCTCTGCC	TAAATTACAT
3721	GTTGGCGTTG	TTAAATATGG	CGATTCTCAA	TTAAGGCCAA	CTGTTGAGCG	TTGGTTTAT

FIG. 7-1

3781	ACTGGTAAGA	ATTTGTATAA	CGCATATGAT	ACTAAACAGG	CTTTTCTAG	TAATTATGAT	3840
3841	TCCGGTGT	ATTCTTATT	AACGCCCTAT	TTATCACACG	GTCGGTATT	CAAACCAATT	3900
3901	AATTAGGTC	AGAAGATGAA	GCTTACTAAA	ATATATTG	AAAAGTTTC	ACGCGTTCTT	3960
3961	TGTCTTGC	TTGGATTTC	ATCAGCATT	ACATATAGT	ATATAACCCA	ACCTAACCG	4020
4021	GAGGTTAAAA	AGGTAGTCTC	TCAGACCTAT	GATTTGATA	AATTCACTAT	TGACTCTTCT	4080
4081	CAGCGTCTT	ATCTAACGTA	TCGCTATGTT	TCAGGGATT	CTAAGGGAAA	ATTAATTAA	4140
4141	AGCGACGATT	TACAGAACGA	AGGTTATTCA	CTCACATATA	TTGATTTATG	TACTGTTTCC	4200
4201	ATTAAGGAAAG	GTAATTCAAA	TGAAATTGTT	AAATGTAATT	AATTGTTT	TCTTGATGTT	4260
4261	TGTTTCTTCA	TCTTCTTTC	CTCAGGTAAT	TGAAATGAAT	AATTGCCCTC	TGCGCGATT	4320
4321	TGTAACCTGG	TATTCAAAGC	AATCAGGCGA	ATCCGTTATT	GTTTCTCCCG	ATGTAAGGAAAG	4380
4381	TACTGTTACT	GTATATTCA	CTGACGTTAA	ACCTGAAAAT	CTACGCAATT	TCTTATTTC	4440
4441	TGTTTACG	GCTAATAATT	TTGATATGGT	TGGTCAATT	CCTTCCATAAA	TTCAGAACGTA	4500
4501	TAATCCAAAC	AATCAGGATT	ATATTGATGA	ATTGCCATCA	TCTGATAATC	AGGAATATGA	4560
4561	TGATAATTCC	GCTCCTTCG	GTGGTTCTT	TGTTCCGCAA	AATGATAATG	TTACTCAAAAC	4620
4621	TTTTAAATT	AATAACGTT	GGGCAAAGGA	TTAATACGA	TGTTGCAAT	TGTTGTAAG	4680
4681	GTCTAACT	TCTAAATCCT	CAAATGTATT	ATCTATTGAC	GGCTCTAAC	TATTAGTTGT	4740
4741	TAGTGCACCT	AAAGATATT	TAGATAACCT	TCCCTCAATT	CTTTCTACTG	TTGATTIGCC	4800
4801	AACTGACCAG	ATATTGATTG	AGGGTTGAT	ATTGAGGTT	CAGCAAGGTG	ATGCTTAAAG	4860
4861	TTTTCATTT	GCTGCTGGCT	CTCAGCGTGG	CACTGTTGCA	GGCGGTGTTA	ATACTGACCG	4920
4921	CCTCACCTCT	TTTTTATCTT	CTGCTGGTGG	TTGTTGGTGT	ATTTTAATG	GCGATGTTT	4980
4981	AGGGCTATCA	GTTCGCGCAT	TAAGACTAA	TAGCCATTCA	AAAATATTGT	CTGTGCCACG	5040
5041	TATTCTTACG	CTTCAGGTC	AGAAGGGTTC	TATCTCTGTT	GGCCAGAAATG	TCCCTTTTAT	5100
5101	TACTGGTCGT	GTGACTGGTG	AATCTGCCAA	TGAAATAAT	CCATTTCAGA	CGATTGAGCG	5160
5161	TCAAAATGTA	GGTATTTC	TGAGCGTTT	TCCTGTTGCA	ATGGCTGGCG	GTAATATIGT	5220
5221	TCTGGATATT	ACCAGCAAGG	CCGATAGTTT	GAGTTCTCT	ACTCAGGCAA	GTGATGTTAT	5280
5281	TACTAATCAA	AGAAGTATTG	CTACAAACGGT	TAATTGCGT	GATGGACAGA	CTCTTTACT	5340
5341	CGGGCCCTC	ACTGATTATA	AAAACACTTC	TCAAGATTCT	GGCCTACCGT	TCCTGTCAA	5400
5401	AATCCCTTA	ATCGGCCTCC	TGTTAGCTC	CCGCTCTGAT	TCCAACGAGG	AAAGCACGTT	5460
5461	ATACGTC	GTCAAAGCAA	CCATAGTACG	CGCCCTGTAG	CGGGCATTAA	AGCGCGGCG	5520
5521	GTGTGGTGGT	TACGCGCAGC	GTGACCGCTA	CACTTGCAG	CGCCCTAGCG	CCCCTCTT	5580
5581	TCGCTTCTT	CCCTTCCTT	CTGCCACG	TCGCCGGCT	TCCCCGTCAA	GCTCTAAATC	5640
5641	GGGGGCTCCC	TTTGGGTT	CGATTAGTG	CTTACGGCA	CCTCGACCCC	AAAAAAACTG	5700
5701	TTTGGGTGA	TGGTTCACGT	AGTGGGCCAT	CGCCCTGATA	GACGGTTTT	CGCCCTTTGA	5760
5761	CGTTGGAGTC	CACGTTCTT	AAATGTTGAC	TCTTGTCCA	AACTGGAACA	ACACTCAACC	5820
5821	CTATCTCGGG	CTATCTTTT	GATTATAAG	GGATTTTGC	GATTTCGGAA	CCACCATCAA	5880
5881	ACAGGATTT	CGCCTGCTGG	GGCAAACCG	CGTGGACCGC	TTGTCGAAAC	TCTCTCAGGG	5940
5941	CCAGGCGGTG	AAGGGCAATC	AGCTGTTGCC	CGTCTCGCTG	GTGAAAAGAA	AAACCACCT	6000
6001	GGGCCAAAT	ACGCAAACCG	CCTCTCCCG	CGCCTGGCC	GATTCATTAA	TGAGCTGGC	6060
6061	ACGACAGGTT	TCCCGACTGG	AAAGCGGGCA	GTGAGCGCAA	CGAATTAAT	GTGAGTTAGC	6120
6121	TCACTCATT	GGCACCCAG	GCTTACACT	TTATGCTCC	GGCTCGTATG	TTGTGTGGAA	6180
6181	TTGTGAGCGG	ATAACAAATT	CACACGCGTC	ACTTGGCACT	GGCCGTCGTT	TTACAACGTC	6240
6241	GTGACTGGGA	AAACCTGGC	TTTACCAAG	CTTGTACAT	GGGAAAATA	AAGTGAACAA	6300
6301	AAGCACTATT	GCACCTGGC	TCTTACCGTT	ACCGTTACTG	TTTACCCCTG	TGACAAAAGC	6360
6361	CGCCCAGGTC	CAGCTGCTG	AGTCAGGCT	ATTGTGCCA	GGGGATTGTA	CTAGTGGATC	6420
6421	CTAGGCTGAA	GGCGATGACC	CTGCTAAGGC	TGCATTCAT	AGTTTACAGG	CAAGTGTAC	6480
6481	TGAGTACATT	GGCTACGCTT	GGGCTATGGT	AGTAGTTATA	GTTGGTGTCA	CCATAGGGAT	6540
6541	TAAATTATTC	AAAAAGTTA	CGAGCAAGGC	TTCTTAAGCA	ATAGCGAAGA	GGCCCGCACC	6600
6601	GATGCCCTT	CCCAACAGT	GCGCAGCCTG	AATGGCGAAT	GGCGCTTTCG	CTGGTTTCCG	6660
6661	GCACCAAGAAG	CGGTGCGGGA	AAGCTGGCTG	GAGTGGCAGTC	TTCTCTGAGGC	CGATACGGTC	6720
6721	GTCGTCCCC	CAAACGGCA	GATGCACGGT	TACGATGCGC	CCATCTACAC	CAACGTAACC	6780
6781	TATCCCATT	CGGTCATCC	GCGTTTGT	CCCACGGAGA	ATCCGACGGG	TTGTTACTG	6840
6841	CTCACATT	ATGTTGATG	AAGCTGGCTA	CAGGAAGGCC	AGACGCCAAT	TATTTTAAACA	6900
6901	GGCGTTCTA	TTGGTTAAA	AATGAGGCTG	TTAACACAAA	ATTAACCGCG	AATTTTAAACA	6960
6961	AAATATTAAAC	TTTACAATT	AAATATTATG	CTTATACAA	CTTCTGT	TTGGGGCTTT	7020
7021	TCTGATTATC	AACCAGGGTA	CATATGATTG	ACATGCTAGT	TTACGATTA	CCGTTCATCG	7080
7081	ATTCCTTGT	TTGCTCAGA	CTCTCAGGCA	ATGACCTGAT	AGCCTTTGTA	GATCTCTCAA	7140
7141	AAATAGCTAC	CCTCTCCGGC	ATTAATTAT	CAGCTAGAAC	GGTTGAATAT	CATATTGATG	7200
7201	GTGATTTGAC	TGTCTCCGGC	CTTCTCACC	CTTGTGAT	TTTACCTACA	CATTACTCAG	7260
7261	GCATTGCAATT	AAAATATAT	GAGGGTTCTA	AAAATTTTA	TCCTTGC	GAAATAAAGG	7320
7321	CTTCTCCCGC	AAAAGTATTA	CAGGGTCATA	ATGTTTTGG	TACAACCGAT	TTAGCTTAT	7380
7381	GCTCTGAGGC	TTTATTGCTT	AATTGGCTA	ATTCTTGCC	TTGCGCTGTAT	GATTATTG	7440
7441	ACGTT	10	20	30	40	50	60

FIG. 7-2

SUBSTITUTE SHEET

11 / 15

	10	20	30	40	50	60	
1	AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTTCAG	CTCGCGCCCC	AAATGAAAAT	60
61	ATAGCTAAAC	AGGTTATTGA	CCATTTCGCA	AATGTATCTA	ATGGTCAAAAC	TAAATCTACT	120
121	CGTTCGCAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCCAGACA	CCGTACTTTA	180
181	GTTGCATATT	AAAAACATGT	TGAGCTACAG	CACCAAGATT	AGCAATTAAAG	CTCTAAGCCA	240
241	TCTGCAAAAAA	TGACCTCTTA	TCAAAAGGAG	CAATTAAAGG	TACTCTCTAA	TCCTGACCTG	300
301	TTGGAGTTTG	CTTCCGGTCT	GGTTCGCTTT	GAAGCTCGAA	TTAAAACGCG	ATATTGAAAG	360
361	TCTTCGGGC	TTCTCTTAA	TCTTTTGT	GCAATCCGCT	TTGCTTCTGA	CTATAATAGT	420
421	CAGGGTAAAG	ACCTGATTTT	TGATTITATG	TCATTTCG	TTTCTGAAC	GTTTAAAGCA	480
481	TTTGAGGGGG	ATTCAATGAA	TATTTATGAC	GATTCCGCAG	TATTGGACGC	TATCCAGTCT	540
541	AAACATTTA	CTATTACCCC	CTCTGGCAAA	ACTTCTTTG	CAAAGGCCTC	TCGCTATTTT	600
601	GGTTTTTATC	GTCGTCGTT	AAACGAGGGT	TATGATAGT	TTGCTCTTAC	TATGCCTCGT	660
661	AATTCTTTT	GGCGTTATGT	ATCTGATTA	GTTGAATGT	GTATTCCTAA	ATCTCAACTG	720
721	ATGAATCTT	CTACCTGTA	TAATGTTGTT	CCGTTAGTTC	GTTTATTAA	CGTAGATTAA	780
781	TCTTCCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTTCTTA	AAATCGCATA	AGGTAATTCA	840
841	CAATGATTA	AGTTGAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCTGGTGT	900
901	CTCGTCAGGG	CAAGCCTAT	TCACTGAATG	AGCAGCTTTG	TTACGTTGAT	TTGGGTAATG	960
961	AATATCCGGT	TCTTGTCAAG	ATTACTCTG	ATGAAGGTCA	GCCAGCCTAT	GCGCTGGTC	1020
1021	TGTACACCGT	TCATCTGTCC	TCTTCAAAAG	TTGGTCAGTT	CGGTTCCCTT	ATGATTGACC	1080
1081	GTCTGCGCCT	CGTTCCGGCT	AAGTAACATG	GAGCAGGTCG	CGGATTTCGA	CACAATTAT	1140
1141	CAGGCATGA	TACAAATCTC	CGTTGACTT	TGTTTCGCGC	TTGGTATAAT	CGCTGGGGT	1200
1201	CAAAGATGAG	TGTTTATGT	TATTCCTTCG	CCTCTTCG	TTAAGGTTGG	TGCCTTCGTA	1260
1261	GTGGCATTAC	GTATTTTAC	CGTTTAATGG	AAACTTCCTC	ATGAAAAAAAGT	CTTTAGTCCT	1320
1321	CAAAGCCTCT	GTAGCCGTTG	CTACCCCTCGT	TCCGATGCTG	TCTTCGCTG	CTGAGGGTGA	1380
1381	CGATCCCGCA	AAAGCGGCT	TTAACTCCCT	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA	1440
1441	TGCGTGGCG	ATGGTTGTTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTTAAGAA	1500
1501	ATTCACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	GGCTCCTTTT	GGAGCCTTTT	1560
1561	TTTTGGAGA	TTTCAACGT	GAAAAAAATT	TTATTGCAA	TTCCCTTGT	TGTTCCCTTC	1620
1621	TATTCTACT	CCGCTGAAAC	TGTTGAAAGT	TGTTGCAA	AACCCCATAC	AGAAAATTCA	1680
1681	TTTACTAACG	TCTGGAAAGA	CGACAAAAACT	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT	1740
1741	CTGTGGAATG	CTACAGGCGT	TGTTAGTTGT	ACTGGTGACG	AAACTCAGTG	TTACGGTACA	1800
1801	TGGGTTCTA	TTGGGCTTC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGCGGT	1860
1861	TCTGAGGGGTG	GCGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGAGTACGG	TGATACACCT	1920
1921	ATTCGGGCT	ATACTTATAT	CAACCCCTCTC	GACGGCACTT	ATCCGCTGG	TACTGAGCAA	1980
1981	AACCCCGCTA	ATCCTAAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCATGTTT	2040
2041	CAGAATAATA	GGTTCCGAAA	TAGGCAGGGG	GCATTAACGT	TTTATACGGG	CACTGTTACT	2100
2101	CAAGGCACTG	ACCCCGTTAA	AACTTATTAC	CAGTACACTC	CTGTATCATC	AAAAGCCATG	2160
2161	TATGACGCTT	ACTGGAACGG	TAATTCAAGA	GACTGCGTT	TCCATTCTGG	CTTAAATGAA	2220
2221	GATCCATTCTG	TTTGTGAATA	TCAAGGCAA	TCGCTGACCC	TGCGCTCAACC	TCCTGTCATT	2280
2281	GCTGGCGGCG	CGTCTGGTGG	TGGGCTGGT	GGCGGCTCG	AGGGGGTGG	CTCTGAGGGT	2340
2341	GGCGGGTCTG	AGGGGCGGCG	CTCTGAGGGA	GGCGGTTCCG	GTGGGGTCTC	TGGGTCGGT	2400
2401	GATTTTGATT	ATGAAAAGAT	GGCAAACGCT	AATAAGGGGG	CTATGACCGA	AAATGCCGAT	2460
2461	GAAAACGCGC	TACAGTCTGA	CGCTAAAGGC	AAACTTGATT	CTGTCGCTAC	TGATTACGGT	2520
2521	GCTGCTATCG	ATGGTTTCAT	TGGTGACGTT	TCCGGCTTG	CTAATGGTAA	TGGTGCTACT	2580
2581	GGTGTATTTG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTACACCT	2640
2641	TTAATGAATA	ATTTCCGTC	ATATTTCACCT	TCCCTCCCTC	AATCGGTTGA	ATGTCGCCCT	2700
2701	TTTGTCTTAA	GCGCTGGTAA	ACCATATGAA	TTTTCTATTG	ATTGTGACAA	AATAAAACTTA	2760
2761	TTCCGTGGTG	TCTTTGCGTT	TCTTTATAT	GTTGCCACCT	TTATGATGT	ATTTTCTACG	2820
2821	TTTGCTAACAA	TACTGCGTA	TAAGGAGTCT	TAATCATGCC	AGTTCTTTTG	GGTATTCCGT	2880
2881	TATTATTGCG	TTTCCCTCGGT	TTCTTCTGTT	TAACTTGTT	CGGCTATCTG	CTTACTTTTC	2940
2941	TTAAAAAGGG	CTTCGGTAAAG	ATAGCTATTG	CTATTCATT	GTTCTTGTCT	CTTATTATTG	3000
3001	GGCTTAACCTC	AAATTCCTG	GGTTATCTCT	CTGATATTAG	CGCTCAATT	CCCTCTGACT	3060
3061	TTGTTCAAGGG	TGTTCAGTTA	ATTCTCCCGT	CTAATGCGCT	TCCCTGTTT	TATGTTATTG	3120
3121	TCTCTGTAAA	GGCTGCTATT	TTCAATTGTTG	ACGTTAACAA	AAAAATCGTT	TCTTATTGTTG	3180
3181	ATTGGGATAA	ATAATATGGC	TGTTTATTGTT	GTAACTGGCA	AATTAGGCTC	TGGAAAGACG	3240
3241	CTCGTTAGCG	TTGGTAAGAT	TTAGGATAAAA	ATTGTAGCTG	GGTGCAAAT	AGCAACTAAT	3300
3301	CTTGATTTAA	GGCTTCAAAA	CCTCCCGCAA	GTCGGGAGGT	TCGCTAAAC	GCCTCGCGTT	3360
3361	CTTAGAATAC	CGGATAAGCC	TTCTATATCT	GATTTGCTTG	CTATTGGCG	CGGTAAATGAT	3420
3421	TCCTACGATG	AAAATAAAAA	CGGCTTGCTT	GTTCTCGATG	AGTGCCTGAC	TTGGTTAAAT	3480
3481	ACCCGTTCTT	GGAATGATAA	GGAAAGACAG	CCGATTATTG	ATTGGTTTCT	ACATGCTCGT	3540
3541	AAATTAGGAT	GGGATATTAT	TTTCTTGTG	CAGGACTTAT	CTATTGTTGA	TAAACAGGCG	3600
3601	CGTTCTGCA	TAGCTGAACA	TGTTGTTTAT	TGTCGTGTC	TGGACAGAA	TACTTACCT	3660
3661	TTTGTGGTGA	CTTTATATTG	TCTTATTACT	GGCTCGAAA	TGCTCTGCGC	TAAATTACAT	3720
3721	GTTGGCGTTG	TTAAATATGG	CGATTCTCAA	TTAAGCCCTA	CTGTTGAGCG	TTGGCTTTAT	3780
3781	ACTGGTAAGA	ATTGTATAA	CGCATATGAT	ACTAAACAGG	CTTTTCTAG	TAATTATGAT	3840

FIG. 8-1

SUBSTITUTE SHEET

15

FIG. 8-2

137 15

	10	20	30	40	50	60
1	AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTTCAG	CTCGCGGCC	AATTGAAAAT
61	ATAGCTAAAC	AGGTTATTGA	CCATTGCGA	AATGTATCTA	ATGGTCAAAAC	TAAATCTACT
121	CGTTCGCAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCCAGACA	CCGTACITTA
181	GTTCATATT	TAACACATGT	TGAGCTACAG	CACAGATTG	AGCAATTAAG	CTCTAAGCCA
241	TCTGCAAAAAA	TGACCTCTTA	TCAAAAGGAG	CAATTAAAGG	TACTCTCTAA	TCCTGACCTG
301	TTGGAGTTTG	CTTCCGGTCT	GGTTCGCTT	GAAGCTCGAA	TTAAAACGCG	ATATTGAAAG
361	TCTTCGGGC	TTCCCTCTAA	TCTTTTGTAT	GCAATCCGCT	TTGCTTCTGA	CTATAATAGT
421	CAGGGTAAAG	ACCTGATTTT	TGATTTATGG	TCATTCTCGT	TTCTGAACG	GTITAAAGCA
481	TTTGAGGGGG	ATTCAATGAA	TATTTATGAC	GATTCCGCG	TATTGGACGC	TATCCAGTCT
541	AAACATTTA	CTATTACCCC	CTCTGGCAAA	ACTTCTTTG	CAAAGGCCTC	TCGCTATTTT
601	GGTTTTATC	GTGCTCTGGT	AAACGAGGGT	TATGATAGTG	TTGCTCTTAC	TATGCTCGT
661	AATTCCCTTT	GGCGTTATGT	ATCTGCATTA	GTTGAATGTG	GTATTCTAA	ATCTCAACTG
721	ATGAATCTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTC	GTTTTATTAA	CGTAGATT
781	TCTTCCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTTCTTA	AAATCGCATA	AGGTAATTCA
841	CAATGATTA	AGTTGAAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCTGGTGT
901	CTCGTCAGGG	CAAGCCTAT	TCACTGAATG	AGCAGCTTG	TTACGTTGAT	TTGGGTAATG
961	AATATCCGGT	TCTTGTCAG	ATTACTCTG	ATGAAGGTCA	GCCAGCCTAT	GCGCCTGGTC
1021	TGTACACCGT	TCATCTGTCC	TCTTCAAAG	TTGGTCAGTT	CGGTTCCCTT	ATGATTGACC
1081	GTCTGCGCCT	CGTTCCGGCT	AAGTAACATG	GAGCAGGTG	CGGATTCGGA	CACAATTAT
1141	CAGGCATGA	TACAAATCTC	CGTTGACTT	TGTTTCGCGC	TTGGTATAAT	CGCTGGGGT
1201	CAAAGATGAG	TGTTTAGTG	TATTCTTCG	CCTCTTCGT	TTAGGTTGG	TGCCTTCGTA
1261	GTGGCATTAC	GTATTTTAC	CGTTTAATGG	AAACTTCCCTC	ATGAAAAAAGT	CTTTAGTCCT
1321	CAAAGCCTCT	GTAGCCGTTG	CTACCCCTCGT	TCCGATGCTG	TCTTCGCTG	CTGAGGGTGA
1381	CGATCCCGCA	AAAGCGGCC	TTAACTCCCT	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA
1441	TGCGTGGCG	ATGGTTGTTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTTAAGAA
1501	ATTCACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	GGCTCCTTTT	GGAGCCTTTT
1561	TTTTGGAGA	TTTCAACGT	GAAAAAAATT	TTATTCGCAA	TTCTTTAGT	TGTTCCCTTC
1621	TATTCTCACT	CCGCTAAC	TGTTGAAAGT	TGTTAGCAA	AACCCCATAC	AGAAAATTCA
1681	TTTACTAACG	TCTGAAAGA	CGACAAAATC	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT
1741	CTGTGGAATG	CTACAGGCGT	TGTTAGTTGT	ACTGGTGACG	AAACTCAGTG	TTACGGTACA
1801	TGGGTTCTA	TTGGGCTGTC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGCGGT
1861	TCTGAGGGTG	GCGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGAGTACGG	TGATACACCT
1921	ATCCGGGCT	ATACTTATAT	CAACCCCTCTC	GACGGCATT	ATCCGCTGG	TACTGAGCAA
1981	AACCCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCATGTTT
2041	CAGAATAATA	GGTCCGAAA	TAGGCAGGGG	GCATTAAC	TTTATACGGG	CACTGTTACT
2101	CAAGGCACTG	ACCCCGTTAA	AACTTATTAC	CACTACACTC	CTGTATCATC	AAAAGCCATG
2161	TATGACGCTT	ACTGGAACGG	TAATTCAAGA	GACTGCGTT	TCCATTCTGG	CTTTAATGAA
2221	GATCCATTG	TTTGTGAATA	TCAAGGCCAA	TCGTCGAC	TGCTCAACC	TCTGTCAAT
2281	GCTGGCGGCG	GCTCTGGTGG	TGGTCTGGT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT
2341	GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGG	GGCGGTTCCG	GTGDTGGCTC	TGGTCCGGT
2401	GATTGGTATT	ATGAAAAGAT	GGCAAAACGCT	AATAAGGGGG	CTATGACCGA	AAATGCCGAT
2461	GAAAACGCGC	TACAGTCTGA	CGCTAAAGGC	AAACTTGATT	CTGTCGCTAC	TGATTACGGT
2521	GCTGCTATCG	ATGGTTTCAT	TGGTGACGTT	TCCGGCTTG	CTAATGGTAA	TGGTGCTACT
2581	GGTGATTTG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTACACT
2641	TTAATGAATA	ATTTCCGTC	ATATTACCT	TCCCTCCCTC	AATCGGTTGA	ATGTCGCCCT
2701	TTTGTCTTTA	GCGCTGGTAA	ACCATATGAA	TTTCTATTG	ATTGTGACAA	AATAAACCTA
2761	TCCTCGGGTG	TCTTTGCGTT	TCTTTATAT	GTTGCCACCT	TTATGTATGT	ATTTCTACG
2821	TTTGTAAACA	TACTGCGTAA	TAAGGAGTCT	TAATCATGCC	AGTTCTTTG	GGTATTCCGT
2881	TATTATTGCG	TTTCCCTCGGT	TTCCCTCTGG	TAACTTGTT	CGGCTATCTG	CTTACTTTTC
2941	TTAAAAAGGG	CTTCGGTAAAG	ATAGCTATTG	CTATTCATT	GTTTCTTGT	CTTATTATTG
3001	GGCTTAAC	AATTCTTG	GGTTATCTC	CTGATATTAG	CGCTCAATT	CCCTCTGACT
3061	TTGTTCAAGGG	TGTTCAAGT	ATTCTCCCGT	CTAATGCGCT	TCCCTGT	TATGTTATT
3121	TCTCTGTA	GGGCTGCTATT	TTCATTTTTG	ACGTTAAACA	AAAAATCGTT	TCTTATTG
3181	ATTGGGATAA	ATATAATGGC	TGTTTATTTT	GTAACTGGCA	AATTAGGCTC	TGGAAAGACG
3241	CTCGTAGCG	TTGGTAAGAT	TCAGGATAAA	ATTGTAAGT	GGTGC	AGCAACTAAT
3301	CTTGATTTAA	GGCTTCAAAA	CCTCCCGCAA	GTCGGGAGGT	TCGCTAAAC	GCCTCGCGTT
3361	CTTAGAATAC	CGGATAAGCC	TTCTATATCT	GATTTGCTT	CTATTGGCG	CGGTAATGAT
3421	TCCTACGATG	AAAATAAAAA	CGGCTTGCTT	GTTCTCGATG	AGTGC	TGGTTAAAT
3481	ACCCGTTCTT	GGAATGATAA	GGAAAGACAG	CCGATTATTG	ATTGGTTCT	ACATGCTCGT
3541	AAATTAGGAT	GGGATATTAT	CTTCCTTGTT	CAGGACTTAT	CTATTGTTGA	TAAACAGGCG
3601	CGTTCTGCA	TAGCTGAACA	TGTTGTTTAT	TGTCGTGTC	TGGACAGAA	TACTTTACCT
3661	TTTGTGCGTA	CTTTATATTC	TCTTATTACT	GGCTCGAAA	TGCTCTG	TAAATTACAT
3721	GTGCGCTTG	TTAAATATGG	CGATTCTCAA	TTAGCCCTA	CTGTTGAGCG	TTGGCTTTAT
3781	ACTGGTAAGA	ATTGTATAA	CGCATATGAT	ACTAAACAGG	CTTTTCTAG	TAATTATGAT

FIG. 9-1

SUBSTITUTE SHEET

15 / 15

3841	TCCGGTGT	TTT	ATTCTTATT	AACGCC	TTATCACAG	GTCGGTATT	CAAACCA	3900
3901	AAITTAGGTC	AGAAGATGAA	GCTTACTAAA	ATATATTG	AAAAGTTTC	ACGCGTTCT	3960	
3961	TGCTTGCGA	TTGGATTG	ATCAGCATT	ACATATAGTT	ATATAACCC	ACCTAAGCCG	4020	
4021	GAGGTTAAAA	AGGTAGTCTC	TCGACCTAT	GATTTTGATA	AATTCACTAT	TGACTCTTCT	4080	
4081	CAGCGTCTTA	ATCTAAGCTA	TCGCTATGTT	TTCAAGGATT	CTAAGGGAAA	ATTAATTAA	4140	
4141	AGCGACGATT	TACAGAAGCA	AGGTTATICA	CTCACATATA	TTGATTATG	TACTGTTTCC	4200	
4201	ATTAAGG	TAATTCAAAT	GAATTGTTA	AATGTAATT	ATTTGT	CTTGATGTT	4260	
4261	GTTCATCAT	CTTCTTTG	TCAGGTATT	GAAATGAATA	ATTGCGCTCT	GCGCATT	4320	
4321	GTAACTTGGT	ATTCAGAAGCA	ATCAGGCAGA	TCCGTTATTG	TTCTCCCGA	TGTAAAGGT	4380	
4381	ACTGTTACTG	TATATTCTAC	TGACGTTAAA	CCTGAAATC	TACGCAATT	CTTATTCT	4440	
4441	GTTCACGTG	CTAATAATT	TGATATGGTT	CCTTCAATT	CTTCCATTAT	TTAGAAGTAT	4500	
4501	AATCCAAACA	ATCAGGATT	TATTGATGAA	TTGCCC	CTGATAATCA	GGAATATGAT	4560	
4561	GATAATTCCG	CTCCTCTG	TGGTTCTTT	GTCAGAAA	ATGATAATGT	TACTCAAAC	4620	
4621	TTTAAATTAA	ATAACGTTCG	GGCAAAGGAT	TTAATACGAG	TTGTCGAATT	GTTCGAAAG	4680	
4681	TCTAATACTT	CTAAATCTC	AAATGTATT	TCTATTGACG	GCTCTAATCT	ATTAGTTGTT	4740	
4741	AGTGCACCTA	AAGATAATT	AGATAACCTT	CCTCAATT	TTTCTACTGT	TGATTGCCA	4800	
4801	ACTGACCAGA	TATTGATTG	GGGTTTGATA	TTTGGAGGTT	AGCAAGGTGA	TGCTTAGAT	4860	
4861	TTTCATTTG	CTGCTGGCTC	TCAGCGTGG	ACTGTTGCG	GCGGTGTTAA	TACTGACC	4920	
4921	CTCACCTCTG	TTTATCTTC	TGCTGGTGGT	TGTTAATGG	CGATGTTTAA	CGATGTTT	4980	
4981	GGGCTATCAG	TTCGCGCATT	AAAGACTAAT	AGCCATTCAA	AAATATTGTC	TGTGCACGT	5040	
5041	ATTCTTACGC	TTTCAGGTCA	GAAGGGTTCT	ATCTCTGTT	GCCAGAATGT	CCCTTTTATT	5100	
5101	ACTGGTCGTG	TGACTGGTGA	ATCTGCCAAT	GTAAATAATC	CATTTCAAGAC	GATTGAGCGT	5160	
5151	CAAATGTAG	GTATTCTCAT	GGCGTTTT	CCTGTTGCAA	TGGCTGGCGG	TAATATTGTT	5220	
5221	CTGGATATT	CCAGCAAGGC	CGATAGTTTG	AGTTCTCTA	CTCAGGCAAG	TGATGTTATT	5280	
5281	ACTAATCAA	GAAGTATTG	TACAACGGTT	AATTGCGT	ATGGACAGAC	TCTTTACTC	5340	
5341	GGTGGCCTCA	CTGATTATAA	AAACACTTCT	CAAGATTCTG	GCGTACCGT	CCTGCTAA	5400	
5401	ATCCCCTTAA	TCGCTCTCT	GTTTAGCTCC	CGCTCTGATT	CCAACAGGAA	AAGCACGTTA	5460	
5461	TA CGT GCT CG	TCAAAGCAAC	CATAGTACGC	GCCCTGTAGC	GCGCATTAA	GCGCGGCGGG	5520	
5521	TGTGGTGGT	ACGCGAGCG	TGACCGTAC	ACTTGGCAG	GCCCTAGCGC	CCGCTCTT	5580	
5581	CGCTTCTTC	CCTCTTTTC	TCGCCACGTT	CGCGGCTTT	CCCCGTCAG	CTCTAAATCG	5640	
5641	GGGGCTCCCT	TTAGGGTTCC	GATTTAGTGC	TTTACGGCAC	CTCGACCCCC	AAAAACTTGA	5700	
5701	TTGGGTGAT	GGTTCACGTA	GTGGGCCATC	GCCCTGATAG	ACGGTTTTTC	GCCCTTGGAC	5760	
5761	GTGGAGTCC	ACGTTCTTA	ATAGTGGAT	CTTGTCCAA	ACTGGAACAA	CACTAACCCC	5820	
5821	TATCTGGGGC	TATTCTTTG	ATTATAAGG	GATTTGCCG	ATTCATTAA	CACCACTAAA	5880	
5881	CAGGATTTC	GCCTGCTGG	GCAAAACAGC	GTGGACCGCT	CTCTAGGGC	CTCTCAGGGC	5940	
5941	CAGGCGGTGA	AGGGCAATCA	GCTGTGCCC	GTCTCGCTGG	TGAAAAGAAA	AACCACCTG	6000	
6001	GCGCCCAATA	CGCAAACCGC	CTCTCCCCGC	GCCTGGCCG	ATTCAATTAT	GCAGCTGGCA	6060	
6061	CGACAGGTT	CCCGACTGGA	AAGCGGGCAG	TGAGCGCAAC	GCAATTAA	TGAGTTAGCT	6120	
6121	CACTCATTAG	GCACCCCCAGG	CTTACACTT	TATGCTTCCG	GCTCGTATGT	TGTGTGGAAT	6180	
6181	TGTGAGCGGA	TAACAATTTC	ACACAGGAAA	CAGCTATGAC	CAGGATGTAC	GAATTGCGAG	6240	
6241	GTAGGAGAGC	TCGGGGATC	CGAGGCTGAA	GGCGATGACC	CTGCTAAGGC	TGCAATTCA	6300	
6301	AGTTTACAGG	CAAGTGCTAC	TGAGTACATT	GGCTACGCTT	GGGTATGGT	AGTAGTTATA	6360	
6361	GTGGTGTCA	CCATAGGGAT	TAAATTATT	AAAAGTTT	CGAGCAAGGC	TTCTTAACCA	6420	
6421	GCTGGCGTAA	TAGCGAAGAG	GCCCGCACCG	ATCGCCCTC	CCAACAGTTG	CGCAGCCTGA	6480	
6481	ATGGCGAATG	GCGCTTGCC	TGGTTCCGG	CACCAAGAAGC	GGTGCCGGAA	AGCTGGCTGG	6540	
6541	AGTGCATCT	TCCTGAGGCC	GATACTGTCG	TCGCCCCCTC	AAACTGGCAG	ATGCACGGTT	6600	
6601	ACGATGCGCC	CATCTACACC	AACGTAACCT	ATCCATTAC	GGTCATCCG	CCGTTGGTTC	6660	
6661	CCACGGAGAA	TCCGACGGGT	TGTTACTCGC	TCACATTAA	TGTTGATGAA	AGCTGGCTAC	6720	
6721	AGGAAGGCCA	GACCGAATT	ATTTTGTATG	GCCTTCCCTAT	TGGTTAAAAA	ATGAGCTGAT	6780	
6781	TTAACAAAAA	TTAACCGCA	ATTTAACAA	AATATTAAACG	TTTACAATT	AAATATTGTC	6840	
6841	TTATACAA	TTCTGTTT	TGGGGCTTT	CTGATTATCA	ACCGGGGTAC	ATATGATTGA	6900	
6901	CATGCTAGTT	TTACGATTAC	CGTTCATCGA	TTCTCTGTT	TGCTCCAGAC	TCTCAGGCAA	6960	
6961	TGACCTGATA	GCCTTGTAG	ATCTCTCAA	AATAGCTACC	CTCTCCGGCA	TTAATTATC	7020	
7021	AGCTAGAACG	GTGAATATC	ATATTGATGG	TGATTGACT	GTCTCCGGCC	TTTCTCACCC	7080	
7081	TTTGTGAT	TTACCTACAC	ATTACTCAGG	CATTGCATT	AAAATATATG	AGGGTTCTAA	7140	
7141	AAATTTTAT	CCTGCGTTG	AAATAAAGGC	TTCTCCCGA	AAAGTATTAC	AGGGTCATAA	7200	
7201	TGTTTTGGT	ACAACCGATT	TAGCTTATG	CTCTGAGGCT	TTATTGCTTA	ATTTTGCTAA	7260	
7261	TTCTTTGCCT	TGCTGTATG	ATTTATTGGA	CGTT			7294	

FIG. 9-2

15/15

	10	20	30	40	50	60
1	AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTTCAG	CTCGCGCCCC	AAATGAAAAT
61	ATAGCTAAAC	AGGTTATTGA	CCATTGCGA	AATGTATCTA	ATGGTCAAAC	TAAATCTACT
121	CGTTCGCAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCCAGACA	CCGTACTTAA
181	GTTGCATATT	TAAAACATGT	TGAGCTACAG	CACCAAGATTC	AGCAATTAAAG	CTCTAAGCCA
241	TCTGCAAAAAA	TGACCTCTTA	TCAAAAGGAG	CAATTAAAGG	TACTCTCTAA	TCCTGACCTG
301	TTGGAGTTG	CTTCCGGTCT	GGTTCGCTT	GAAGCTCGAA	TTAAAACGCG	ATATTGAAAG
361	TCTTTCGGGC	TTCCCTTAA	TCTTTTGT	GCAATCCGCT	TTCGTTCTGA	CTATAAATAGT
421	CAGGGTAAAG	ACCTGATTT	TGATTTATGG	TCATTCTCGT	TTTCTGAACT	GTTTAAAGCA
481	TTTGGGGGG	ATTCAATGAA	TATTTATGAC	GATTCCGCGAG	TATTGGACGC	TATCCAGTCT
541	AAACATTTTA	CTATTACCC	CTCTGGCAAA	ACTTCTTTG	CAAAGCCTC	TCGCTATTTT
601	GGTTTTTATC	GTCGTCGTT	AAACGAGGGT	TATGATAGTG	TIGCTCTTAC	TATGCTCGT
661	AATTCCCTTT	GGCGTTATGT	ATCTGCATTA	GTTGAATGTG	GTATTCCTAA	ATCTCAACTG
721	ATGAATCTTT	CTACCTGTAA	TAATGTTGTT	CGGTTAGTTC	GTTTATTAA	CGTAGATTTT
781	TCTTCCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTTCTTA	AAATCGCATA	AGGTAATTCA
841	CAATGATTAA	AGTTGAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCTGGTGT
901	CTCGTCAGGG	CAAGCCATTAT	TCACTGAATG	AGCAGCITTG	TTACGTTGAT	TTGGGTAATG
961	AATATCCGGT	TCTTGTCAGG	ATTACTCTTG	ATGAAGGTCA	GCAGCCTAT	GCGCCTGGTC
1021	TGTACACCGT	TCATCTGTCC	TCTTCAAAG	TTGGTCAGTT	CGGTTCCCT	ATGATTGACC
1081	GTCGCGCT	CGTTCGGCT	AAGTAACATG	GAGCAGGTCG	CGGATTTGCA	CACAATTAT
1141	CAGGGCGATGA	TACAAATCTC	CGTTGACTT	TGTTTGCGC	TGGTATAAT	CGCTGGGGGT
1201	CAAAGATGAG	TGTTTATGT	TATTCTTCG	CCTCTTCTCGT	TTAGGTTGG	TGCTTCGTA
1261	GTGGCATTAC	GTATTTTAC	CGTTTAATGG	AAACTTCTC	ATGAAAAAAGT	CTTAGTCTT
1321	CAAAGCCTCT	GTAGCCGTTG	CTACCCCTCGT	TCCGATGCTG	TCTTCGCTG	CTGAGGGTGA
1381	CGATCCCGCA	AAAGCGGCCT	TTAACTCCCT	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA
1441	TGCGTGGGCG	ATGGTTGTTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTTAAGAA
1501	ATTCACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	GGCTCCTTTT	GGAGCCTTTT
1561	TTTTGGAGA	TTTCAACGT	GAAAAAATTA	TTATTGCAA	TTCTTTAGT	TGTTCCCTTC
1621	TATTCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTAGCAA	AACCCCATAAC	AGAAAATTCA
1681	TTTACTAACG	TCTGAAAGA	GCACAAAATC	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT
1741	CTGTGGAATG	CTACAGGCCT	TGTAGTTGTT	ACTGGTGACG	AAACTCAGTG	TTACGGTACA
1801	TGGGTTCTA	TTGGGCTTC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGCGGT
1861	TCTGAGGGTG	GCGGTTCTGA	GGGGCGCGGT	ACTAAACCTC	CTGAGTACGG	TGATACACCT
1921	ATTCGGGCT	ATACTTATAT	CAACCTCTC	GACGGCACTT	ATCCGCTCTGG	TACTGAGCAA
1981	AACCCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAAATAC	TTTCATGTTT
2041	CAAGATAATA	GGTCCGAAA	TAGGCAGGGG	GCATTAACGT	TTTATACGGG	CACTGTTACT
2101	CAAGGCACTG	ACCCCGTTAA	AACTTATTAC	CAGTACACTC	CTGTATCATC	AAAAGCCATG
2161	TATGACGCTT	ACTGGAACCG	TAATTTCAGA	GACTGCGCTT	TCCATTCTGG	CTTTAATGAA
2221	GATCCATTG	TTTGTGAATA	TCAAGGCCAA	TCGTCTGACC	TGCTCTAAC	TCTGTCAAT
2281	GCTGGCGGCG	GCTCTGGTGG	TGGTTCTGGT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT
2341	GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGG	GGCGGTTCCG	GTGGTGGCTC	TGGTTCCGGT
2401	GATTTGATT	ATGAAAAGAT	GGCAAACGCT	AATAAGGGGG	CTATGACCGA	AAATGCCGAT
2461	GAAAACGCGC	TACAGTCTGA	CGCTAAAGGC	AAACTTGATT	CTGTCGCTAC	TGATTACGGT
2521	GCTGCTATCG	ATGGTTTCAT	TGGTGACGTT	TCCGGCTCTG	CTAATGGTAA	TGGTGCTACT
2581	GGTATTGTTG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTACACCT
2641	TTAATGAATA	ATTTCCGTTA	ATATTACCT	TCCCTCCCTC	AATCGGTTGA	ATGTCGCCCT
2701	TTTGTCTTA	GCGCTGGTAA	ACCATATGAA	TTTCTATTG	ATTGTCACAA	AATAAACCTA
2761	TTCCGTGGT	TCTTTCGTT	TCTTTATAT	GTTGCCACCT	TTATGTTATGT	ATTTCTACG
2821	TTTGCTAAC	TACTGCGTA	TAAGGAGTCT	TAATCATGCC	AGTTCTTTG	GGTATTCCGT
2881	TATTATTGCG	TTTCCCTGGT	TTCTTCTGG	TAACTTTGTT	CGGCTATCTG	CTTACTTTTC
2941	TTAAAAAGGG	CTTCGGTAAG	ATAGCTATTG	CTATTTCAT	TTTCTTGCT	CTTATTATTG
3001	GGCTTAACTC	AATTCTTG	GGTTATCTCT	CTGATATTAG	CGCTCAATT	CCCTCTGACT
3061	TTGTTCAAGGG	TGTTCAAGTA	ATTCTCCCGT	CTAATGCGCT	TCCCTGTTT	TATGTTATT
3121	TCTCTGTAAA	GGCTGCTATT	TTCATTTTG	ACGTTAAACA	AAAAATCGTT	TCTTATTGG
3181	ATTGGGATAAA	ATAATATGGC	TGTTTATTTT	GTAACTGGCA	AATTAGGCTC	TGGAAAGACG
3241	CTCGTTAGCG	TTGGTAAGAT	TTAGGATAAA	ATTGTAGCTG	GGTGCAAAAT	AGCAACTAAT
3301	CTTGATTAA	GGCTTCAAAA	CCTCCCGCAA	GTCGGGAGGT	TCGCTAAAC	GCCTCGCGTT
3361	CTTAGAATAC	CGGATAAGCC	TTCTATATCT	GATTTGCTTG	CTATTGGCG	CGGTAATGAT
3421	TCCTACGATG	AAAATAAAAAA	CGGCTTGTCT	GTTCTCGATG	AGTGCCTGAC	TTGGTTTAAT
3481	ACCCGTTCTT	GGAAATGATAA	GGAAAGACAG	CCGATTATG	ATTGGTTCT	ACATGCTCGT
3541	AAATTAGGAT	GGGATATTAT	TTTTCTTGTT	CAGGACTTAT	CTATTGTTGA	TAAACAGGCG
3601	CGTTCTGCAT	TAGCTGAACA	TGTTGTTAT	TGTCGTGTC	TGGACAGAAAT	TACTTACCT
3661	TTTGTGCGTA	CTTTATATT	TCTTATTACT	GGCTCGAAA	TGCTCTGCC	TAAATTACAT
3721	GTTGGCGTTG	TTAAATATGG	CGATTCTCAA	TTAAGCCCTA	CTGGTGGAGCG	TTGGCTTTAT
3781	ACTGGTAAGA	ATTGTATAAA	CGCATATGAT	ACTAAACAGG	C.TTTTCTAG	TAATTATGAT

FIG. 10-1

3841	TCCGGTGT	TTT	ATTCTT	TTT	AACGCCTT	AT	TTATCACACG	GTCGGT	TATTT	CAAACCATT	3900
3901	AATTTAGGT	C	AGAAGATG	AA	GCTTACTAA	AA	ATATATTG	AAAAGTTTC	ACCGT	TCTT	3960
3961	TGTCTTGC	GA	TTGGATT	TC	ATCAGCATT	AC	ATATAGTT	ATATAACCC	ACCTAAGCCG	4020	
4021	GAGGTT	AAAAA	AGGTAGT	CTC	TCAGACCTAT	GAT	TTGATA	AATTCACTAT	TGACT	TCTTCT	4080
4081	CAGCGT	CTT	ATCTAAG	CTA	TGCTATGTT	TTCAAGG	TTGAT	CTAAGGGAAA	ATTAA	TTAAT	4140
4141	AGCGACG	GATT	TACAGAAG	CA	AGGTATTCA	CTCAC	TTGATT	TTGATTTATG	TACTG	TTTCC	4200
4201	ATTA	AAAAAAG	GTAATT	CAA	TGAAATTGTT	AAATG	AA	AATTTGTTT	TCTT	GATGTT	4260
4261	TGTTT	CATCA	TCTT	TTT	TCAGGTAAT	TGA	AA	AATTGCGCTT	TGCGC	GATTT	4320
4321	TGTAAC	TTG	TATT	AAAGC	ATCAGGCGA	ATC	GG	TTTCTCCG	ATG	AAAAGG	4380
4381	TACTG	TACT	GTAT	ATT	CTGACGTTAA	AC	CTG	CTTTT	TCTT	TATTC	4440
4441	TGTTT	TACGT	GCTAA	ATT	TTGATATGTT	TG	GGT	CTAATT	CTAC	GAAAGT	4500
4501	TAATCC	AAAC	AATCAGG	GATT	ATATTGATG	AT	TTG	GATAATC	TTACT	CAAAC	4560
4561	TGATA	ATTCC	GCTC	TTCTG	GTGGTT	TT	TAATACG	GTTG	TGAAT	TGTTG	4680
4621	TTT	AAAATT	AATAACG	TTT	GGGCAAAGGA	TT	ATCTATGAC	GGCT	CTAATC	TATTAG	4740
4681	GTCTAA	TACT	TCTAA	ATCCT	CAAATG	TT	GGT	CTACTG	TTGATT	TGCC	4800
4741	TAGTG	CACT	AAAGA	TATT	TAGATAA	AC	GGG	GAGGTT	ATG	CTT	4860
4801	AACTG	ACCA	ATATT	GATTG	AGGG	TT	CTAG	TGCA	GGCG	GTGTT	4920
4861	TTTT	CATT	GCTG	TGGCT	CTCAG	GG	TGCT	GGTGG	ATTTT	TAATG	4980
4921	CCT	CACCT	GTTT	ATCCT	CTG	GG	TTCG	TGGGT	GCGAT	GTTTT	5040
4981	AGGG	CTATCA	GTT	CGCG	TAAAG	ACTA	AAAG	ACTAA	CTG	TGCCACG	5100
5041	TATT	CTTACG	CTT	TCAGG	CTG	AGA	AGGG	GT	CC	AGAAATG	5160
5101	TACT	GGT	CGT	GT	GACT	GGG	AT	CTG	CTG	CGATTGAGCG	5220
5161	TCAAA	ATG	TA	GGT	TAT	CC	GCG	TGCA	ATG	ATTG	5280
5221	TCTGG	ATAT	ACCAG	CAAGG	CCG	ATAG	GAG	GAGT	ACTC	AGGAA	5340
5281	TACTA	ATCAA	AGAAG	TATTG	CTAC	AAC	GGG	GGT	GGC	TCTTT	5400
5341	CGGT	GGCCTC	ACTG	ATTATA	AA	AAAC	ACTC	AAAG	CC	ACGTT	5460
5401	AATCC	CTT	ATC	GGC	CTCC	TG	TTAG	CTG	CG	GGCG	5520
5461	ATACG	TGTC	GTCAA	AAAGC	CC	ATAG	TACG	GGG	CG	CGCC	5580
5521	GTGTGG	TGTT	TACG	CGC	GTG	ACCG	CTG	GGG	TCC	TGTCTAA	5640
5581	TCG	CTT	CC	CTT	CTG	CTG	CG	GG	CG	TGACCCC	5700
5641	GGGGG	CTCCC	TTT	AGGG	TT	CG	ATT	GGG	CT	AAAAACTTG	5760
5701	ATTTGGG	TGA	TGG	TACG	AG	GGG	GGG	GGG	GG	CGCC	5820
5761	CGTTGG	GAGTC	CACG	TCTT	AAT	AGT	GGG	GGG	GG	GGG	5880
5821	CTAT	CTCGGG	CTAT	CTT	GATT	TATA	GGG	GGG	GG	GGG	5940
5881	ACAGG	ATT	CGC	CTG	GGC	AA	AGT	GGG	GG	GGG	6000
5941	CCAGG	CGGTG	AAGG	CAATC	AG	CTG	GGG	GGG	GG	GGG	6060
6001	GGCG	CCCAAT	ACGG	AAACCG	CCT	CTCCC	GGG	GGG	GG	GGG	6120
6061	ACGAC	AGGTT	TCCC	GA	AAAG	CGGG	GGG	GGG	GG	GGG	6180
6121	TCA	CTT	GG	GG	TAC	GGG	GGG	GGG	GG	GGG	6240
6181	TTGTG	AGCGG	ATAAC	AAATT	GGT	ACCA	GGG	GGG	GG	GGG	6300
6241	GTG	ACTGGG	AAAC	CTGG	TCTT	ACCG	GGG	GGG	GG	GGG	6360
6301	AAG	CACTATT	GCAC	TGG	GG	GGG	GGG	GGG	GG	GGG	6420
6361	GAGG	CATCCG	GGAG	GCTG	GGG	GGG	GGG	GGG	GG	GGG	6480
6421	AAAG	TGCTACT	GAG	TAC	GCT	ACGG	GGG	GGG	GG	GGG	6540
6481	CATAGGG	GATT	AAATT	ATCA	AA	AGT	GGG	GGG	GG	GGG	6600
6541	GCCC	GCACCG	ATCG	CCCT	AA	ACG	GGG	GGG	GG	GGG	6660
6601	TGG	TTCCGG	CAC	CAGAAGC	GG	GGG	GGG	GGG	GG	GGG	6720
6661	GATAC	GTG	TG	CTCCC	TT	GG	GGG	GGG	GG	GGG	6780
6721	AACTG	TAACCT	ATCC	ATTAC	AA	ACT	GGG	GGG	GG	GGG	6840
6781	TGTTA	CTCG	TCA	ATTAA	TG	GG	GG	GG	GG	GGG	6900
6841	ATTTT	GATG	GC	TTCT	AA	GGG	GGG	GGG	GG	GGG	6960
6901	ATTTT	AAACAA	AA	TATTAACG	TT	GGG	GGG	GGG	GG	GGG	7020
6961	TGGGG	CTTTT	CTG	ATTATCA	AC	GGG	GGG	GGG	GG	GGG	7080
7021	CGTT	CATCGA	TTCT	CTTGTG	TG	GG	GG	GGG	GG	GGG	7140
7081	ATCT	CTCAA	AA	TAGCTAC	CT	CTC	GGG	GGG	GG	GGG	7200
7141	ATATTG	ATG	TG	TTTGACT	GT	CTC	GGG	GGG	GG	GGG	7260
7201	ATTACT	CAGG	CATT	GCATT	AA	GGG	GGG	GGG	GG	GGG	7320
7261	AAATAAAGG	C	TTCT	CCCGCA	AAAG	TATTAC	GGG	GGG	GG	GGG	7380
7321	TAG	TTTATG	CTCT	GAGGCT	TT	ATTG	GCTTAA	TTT	GGG	GGG	7394
7381	ATTT	ATTG	GG	CGTT	10	20	30	40	50	60	

FIG. 10-2

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US91/07141

I. CLASSIFICATION OF SUBJECT MATTER (If several classification symbols apply, indicate all) ⁶

According to International Patent Classification (IPC) or to both National Classification and IPC
 IPC(5): C12N 1/24, 15/00; C07H 21/00
 U.S. CL.: 435/252.33, 320.1, 172.3; 536/27

II. FIELDS SEARCHED

Minimum Documentation Searched ⁷

Classification System ⁸	Classification Symbols
U.S.	435/252.33, 320.1, 172.3, 69.1; 536/27
Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched ⁹	

APS, CAS: search terms: Codon bins, codon preference

III. DOCUMENTS CONSIDERED TO BE RELEVANT ¹⁰

Category ¹¹	Citation of Document, ¹² with indication, where appropriate, of the relevant passages ¹³	Relevant to Claim No. ¹⁴
Y	EP. A. 0.383.620 (Cook) 22 August 1990. See entire document.	1-87
Y	US. A. 4,458,066 (Caruthers et al.) 03 July 1984. see entire document.	1-87
Y	US. A. 4,771,000 (Vertipps et al.) 13 September 1988. see entire document.	8, 9, 24-26 32-34. 55-57. 64-66. 73-75. 81-87
Y	APPLIED MICROBIOLOGY AND BIOTECHNOLOGY. Volume 21. issued 1985. J.M. Jaynes et al., "Construction and expression of synthetic DNA fragments coding for polypeptides with elevated levels of essential amino acids". pages 200-205. see entire document.	1-87

* Special categories of cited documents ¹⁰

- A" document defining the general state of the art which is not considered to be of particular relevance
- E" earlier document but published on or after the international filing date
- L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- O" document referring to an oral disclosure, use, exhibition or other means
- P" document published prior to the international filing date but later than the priority date claimed

-T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

-X" document of particular relevance. the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

-Y" document of particular relevance. the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

-A" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search

11 December 1991

Date of Mailing of this International Search Report

22 JAN 1992

International Searching Authority

ISA/US

Signature of Authorized Officer

James Ketter

ebw

III DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category	Citation of Document, with indication where appropriate, of the relevant passages	Relevant to Claim No
Y	GENE, Volume 44, issued 1986. A.R. Oliphant, "Cloning of random-sequence oligodeoxynucleotides", pages 177-183. see entire document.	1-87
Y	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCE, Volume 87, issued August 1990, Gwirla et al., "Peptides on phage: A vast library of peptides for identifying ligands", pages 6378-6382. see entire document.	1-87
Y	SCIENCE, Volume 249, issued 27 July 1990, J.J. Devlin, "Random Peptide Libraries: A Source of Specific Protein Binding Molecules", pages 404-406. see entire document.	1-87
Y	SCIENCE, Volume 249, issued 27 July 1990, J.K. Scott, "Searching for Peptide Ligands with an Epitope Library", pages 386-390. see entire document.	1-87
Y	E.-L. WINNACKER, "From Genes to Clones: Introduction to Gene Technology", published 1987 by VCH VmbH (Weinheim, Germany), See pages 276-279, especially Table 7-4.	1-87
Y	SCIENCE, Volume 228, issued 14 June 1985, G.P. Smith, "Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface", pages 1315-1317. see entire document.	8.9.24-26. 32-34. 55-57. 64-66. 73-75. 81-87